# Statistical Group Comparison

# 分组比较的统计分析

廖福挺　著
高　勇　译

重慶大學出版社

**图书在版编目(CIP)数据**

分组比较的统计分析/(美)廖福挺(Liao,T. F.)著;
高勇译. —重庆:重庆大学出版社,2007.3
(万卷方法. 社会科学研究方法经典译丛)
书名原文:Statistical Group Comparison
ISBN 978-7-5624-3942-4
Ⅰ.分… Ⅱ.①廖…②高… Ⅲ.统计分析 Ⅳ.C813
中国版本图书馆 CIP 数据核字(2007)第 015357 号

**分组比较的统计分析**
廖福挺 著
高 勇 译
沈崇麟 审校
责任编辑:雷少波 李定群 版式设计:雷少波
责任校对:任卓惠 责任印制:张 策
*
重庆大学出版社出版发行
出版人:张鸽盛
社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内
邮编:400030
电话:(023) 65102378 65105781
传真:(023) 65103686 65105565
网址:http://www.cqup.com.cn
邮箱:fxk@cqup.com.cn (市场营销部)
全国新华书店经销
重庆华林天美印务有限公司印刷
*
开本:787×1092 1/16 印张:13.25 字数:251千
2007 年 4 月第 1 版 2007 年 4 月第 1 次印刷
印数:1—3 000
ISBN 978-7-5624-3942-4 定价:35.00 元

---

# 作译者简介

**廖福挺**,美国伊利诺大学香槟分校(University of Illinois at Urbana-Champaign)社会学和统计学教授,社会学系主任。

**沈崇麟**,中国社会科学院社会学研究所研究员。

**高勇**,北京市社会科学院助理研究员。中国社会科学院研究生院社会学硕士,北京大学社会学博士。主要研究领域为社会分层与流动、社会学研究方法。

# 总序　社会研究方法的现状及其发展趋势

近年来,社会调查技术和社会研究方法都有很大的发展。在调查技术方面,自20世纪70年代以来,社会变迁多次横断面的跟踪调查研究,几乎成为所有国家和地区了解社会结构转变和社会发展状况的基础性调查。这种调查不仅对社会学的研究有很大促进,对整个社会科学的研究都产生了重大影响,而且这些调查结果有的已成为政府有关部门决策的重要依据。国际上比较著名的此类调查有:美国芝加哥大学全国民意调查中心(National Opinion Research Center,简称 NORC)的“社会综合调查(General Social Survey ,简称 GSS)”,英国埃塞克斯大学调查中心进行的“全国家庭生活和社会变迁调查”,法国经济和社会调查所进行的“全国经济社会调查”,日本社会学会组织进行的“全国社会分层与社会流动调查(简称 SSM)”。中国台湾“中央”研究院社会学研究所,也每两年进行一次“台湾社会变迁基本调查”。美国的“社会基础调查”,现在已成为年度性的调查项目,它是美国国家基金会目前资助的最大的社会科学研究项目。以上这些调查,除美国的调查外,一般均因经费原因采用纵向的间隔性重复调查法,即每隔一段时间,进行一次全国规模的抽样调查。每次调查除保留社会研究所需的基本项目外,每次都有不同的主题。在间隔若干时间后,再重复同一主题的调查,这样的研究设计,使社会变迁研究在可以涉及更为广泛的研究领域的同时,具有更好的积累性和可比性。多年来,这些基础性调查获得的资料,滋养着大批的社会科学研究者,有时一项调查就有很多名博士生用来写博士论文,以此所取得的研究成就,其可靠性受到社会科学界的广泛认同。例如1997年出版的,以台湾社会变迁基本调查数据为基础的研究报告集《90年代的台湾社会,社会变迁基本调查研究系列二》收论文16篇,内容涉及社会生活的各个方面,在台湾引起了极大的反响。

国内社会科学界在这方面也有了长足的发展。笔者所在的中国社会科学院社会学研究所的社会调查和方法研究室,组织或参与了多项与社会变迁有关的大规模抽样调查,取得了一定的研究成果,并积累了大量有关社会变迁的宝贵数据资料,其中主要有:

1. 城乡家庭变迁系列调查:该课题是由中国社会科学院社会学研究所牵头,联合北京大学和地方社科院的研究人员展开的一项类似多次横断面的城乡家庭变迁调查。这一调查始于1981年的“中国五城市婚姻家庭调查”,而后有1988年的“中国农村家庭调查”、1991年的“中国七城市家庭调查”、1998年的“中国城乡家庭变迁调查”。

2. 有关中国城乡社会变迁的系列调查：调查始于1991年的第二批国情调查，然后有1992年的“中国城乡居民生活调查”、1993年的“第三批国情调查”、1995年的“第四批国情调查”和1997年的“中国沿海发达地区社会变迁调查”。上述调查虽然还不是严格意义上的多次横断面的纵贯研究，但研究者已在研究设计中尽量考虑到纵贯研究的基本原则，如调查队伍的稳定、指标的可比性和样本空间的延续性等。

3. 中国城乡社会变迁调查：这一调查开始于2000年，为中国社会科学院重大课题。目前已经完成第一期第一次调查和第二次调查，今后将把这一调查发展为连续的、定期进行的社会变迁调查。

在纵向调查技术取得长足进步的同时，20世纪末至今，电话调查技术也有很大发展。电话调查涉及的范围几乎与个别（面对面）访谈同样全面。电话调查中使用的一系列方法，是在20世纪70年代后期和面对面调查一起发展起来的。在20世纪80年代中，电话调查开始变得很普遍，且成为许多场合中各种调查方法的首选。正如某些学者所言，一种在公共和私营部门被人们用来帮助提高决策效率的收集信息的有效方法为人们所普遍认同时，这一现象本身就具有方法论上的意义。不仅如此，电话调查还有很大的实践意义，因为它为研究者提供了更多的控制调查质量的机会。这一机会包括抽样、被调查人的选择、问卷题项的提问、计算机辅助电话访谈（CATI）和数据录入。正因为如此，今天在各种社会调查中，如果没有发现其他重要的足以放弃使用电话调查的原因，电话调查由于其独特的对调查质量进行全面监控的优点，常常成为各种调查方式的首选。由笔者翻译，重庆大学出版社出版的《电话调查方法：抽样、选择和督导》一书，也于2005年面世。

无论是纵向调查抑或电话调查，实际上都是收集研究资料的方法，而应用社会科学的发展，不仅在于调查技术，即收集资料技术的发展，还在于研究方法和分析技术的发展。近年来，无论是定性研究方法，还是定量研究方法都有了长足的发展。

首先，计算机技术的发展可谓突飞猛进，它对当今社会生活的各个方面产生了巨大的影响，在悄悄地改变着社会科学的研究风格和研究方式的同时，也大大提升了社会科学学者的研究能力。这种影响表现在研究过程的各个阶段，从理论建构（概念映射）、问卷设计（专业的问卷设计软件）、调查实施（计算机辅助访谈、计算机辅助电话访问系统、网络在线调查系统）、数据录入（光学标记识别软件）到数据分析（包括文本、声音、图像资料的处理），甚至延伸到写作发表阶段。这样的过程发生在如社会学、经济学、政治学、心理学、教育学中，促进了学科之间的相互借鉴和交叉融合，至少在研究方法上呈现这种趋势。随着计算机计算能力的大幅度提高，20世纪80年代之后，统计学领域内发生了一场“革命”，主要表现在对定类和定序变量的建模能力的大幅度提高上，以及与分布无关的统计分析模型的发展之上，特别是基于“Resampling”（包括Bootstrap、Jackknife、Monte Carlo模拟等）的建模

技术*。同时,计算能力的提高还带动了基于神经网络、动态模拟、人工智能、生态进化等新兴的分析和预测模型的发展。这些进展都为定量社会科学研究提供了更多的可供选择的工具。

亚德瑞安·E·拉夫特里(Adrian E. Raftery)依据社会学家所处理的数据类型,将定量社会学在美国的发展划分为三个时代:第一代起始于20世纪40年代,交互表是其主要处理对象,研究重点是关联度和对数线性模型;第二代起始于20世纪60年代,主要处理单层次的调查数据,Lisrel类型的因果模型和事件史分析是其研究重点;第三代起始于20世纪80年代后期,开始处理诸如文本、空间、社会网络等非传统的数据类型,目前尚没有形成成熟的形态。拉夫特里的综述,虽然更强调定量社会学研究对统计学的贡献,但也大致勾勒出定量社会学在国外的发展脉络。

从分析模型的角度来看,定量分析在以下几个方向有了突破性发展:

1.缺失值处理:由于社会生活的复杂性,社会调查数据常常出现缺失值,传统的处理方式是忽略这些缺失值,或者用均值替代。但现在则倾向于用多重插值法(multiple implation)或者其他基于模型的方法进行处理。这些技术的发展,不仅会增强我们对数据的处理能力,而且将改变我们设计问卷的方式。基于这些技术,我们在不增加被访者负担的前提下,大大增加调查问卷的内容:每个被访者只回答问卷的一部分,然后通过对缺失值的处理,获得他们对未回答部分的估值。

2.非线性关系:线性假定是经典定量分析的一个常见假定,但在实际研究当中,线性假定只能被看作是对社会现实的一个逼近和简化。面对具体的研究数据,如果没有理论上的明确指引(不幸的是,我们常常没有中程理论的指引),我们是无法在线性模型和非线性模型之间做出取舍的。但MARS模型的出现,让我们可以从经验数据当中获得最为拟合的变量之间的函数关系,而不必预先做出线性假定。这样,理论思考和数据分析就可以实现一个互动的循环过程,定量分析就不单单是对理论和假设的简单证伪过程,而是理论思维一个重要组成部分。

3.测量层次:20世纪六七十年代的统计模型,大多要求数据的测量层次在定距以上,如因素分析,但社会学的调查数据却大多为定类或定序数据。对应分析、Loglinear、Logit、Logistic Regression、潜类分析、Ordinal Regression、Normal Ogive Regression等统计模型的出现,大大提高了定量社会学处理定类和定序数据的能力。

4.测量模型:基于文化、社会、心理和认知等方面的考虑,在社会学界仍有人对问卷调查在中国的效度提出质疑。抛弃“本土化”的文化执著,我们更应当关注的是问卷调查的项目反应理论(item response theory),即被访者回答问卷题器时的过程模型。这方面的进展主要表现在两个方面:一是分解测量量表的成分,如Rasch model、IRT分析、Mokken分析等;二是将测量模型与因果模型或其他分析模型结合在一起,明确把测量误差引入到分析当中,充分评估它们对分析结果的影响,如结

* 对于当前一些国内尚无公认译法的模型、软件等的名称,本“译丛”都只给出了英文的表述,以免造成称谓上的混乱。

构方程模型。

5. 潜变量模型：与测量模型相关联的另外一个发展方向是潜变量模型，例如，潜变量分层分析（latent class analysis）、潜变量结构分析（latent structure analysis）、潜变量赋值分析（latent budget analysis）等。“潜变量”这一概念表明，我们可以通过测量“显变量”来测量无法直接观察的理论概念，如权力、声望、地位等。这样，理论和现实之间，通过“潜变量”到“显变量”的映射（测量过程），就有了连接的桥梁。

6. 分析单元的层序性：在定量分析当中，我们常常强调要避免出现“生态谬误”，即分析单元的层次和结论或推论的层次不一致。与其相关的方法论争论是“宏观和微观”的问题。随着多层次模型的出现，我们可以同时考察多个层次上的问题，我们可以把个人放在其家庭背景中，再把家庭放在社区的背景下，考察个人层次的变量对社区变量的效应，或者社区层次的变量对个体行为的具体影响。在定量分析模型当中，“宏观和微观”的联接获得了建模技术上的支持。在这个领域当中，还有一个方向也值得关注：分析宏观层次的数据，对微观层次进行推论。

7. 社会网络模型：区分“关系数据”和“属性数据”，是把分析重点从个体/群体等社会单元转移到这些社会单元之间关系的第一步，社会网络模型是目前发展较快的一个定量分析领域，其理论根基是结构主义。社会网络分析目前仍然具有较浓厚的“形态学”特征（基于图论的缘故），但却为我们理解社会关系在社会空间上的形态奠定了基础，通过计算机模拟和研究社会网络的历期数据，研究社会结构的“发生学”性质模型也处在萌芽状态当中。

8. 系统动力学：如果说社会网络模型是在社会空间上拓展定量社会学的研究手段，那么，社会过程在时间上和物理空间上的属性，则是事件史模型、事件数模型、历期分析、Cox 回归、时间序列分析、Cohort 分析、状态空间模型等模型的研究对象。在这个领域，计量经济学为定量社会学研究提供了许多有益的范例。

9. 预测模型：上述模型仍然是在分析主义的范式下。有些社会学的应用研究，更强调模型的预测精度，而不是模型的认知价值，例如，社会趋势的预测。由于计算能力的提高，神经网络、基因算法、人工智能、模式识别等数据挖掘技术有了长足发展，已经出现了许多拟合经验数据的预测模型，比较成功的应用出现在计量经济学领域（如对股市的预测）。

10. 计算机模拟：对于社会学应用研究而言，研究的对象具有历史性、规模大、变迁的过程不仅漫长且表现某种渐进性，且因社会隔离/社会伦理原因无法接近或有实验禁忌等，无法直接进行观察和研究，这时计算机模拟就成为一个可供选择的替代方案。计算机模拟主要有两个类型，一是基于计算机网络的模拟：每台微机作为一个代理，整个网络作为“社会”，实时演化，如法国的 Swarm 计划；二是基于概念模型的系统，在计算机时间上，按照既定规则运行，较有名的研究是罗马俱乐部的《增长的极限》，常见的软件有 Simul，Arena 等。自然科学家对此方向似乎比社会学家更有兴趣。

定性研究方法一直是社会学研究领域中比较传统的研究方法，在社会学研究的古典时期，它甚至是社会学家手中唯一的研究方法。但随着定量研究方法在社

会学研究中的广泛应用,定性研究方法就似乎越来越不受到人们的重视。但需要澄清的事实是,在定量分析模型取得飞速发展的同时,在过去的二十多年里,定性研究方法也有了长足的进步。主要表现在以下六个方面:

1.研究素材日益扩大:除了传统的参与观察、深度访谈、专题小组访谈之外,会话、交谈、电视、广播、文档、日记、叙事、自传(autobiography)等社会过程中自然产生的素材,甚至社会学理论本身(理论的形式化),也开始进入定性分析的视野当中。所有这些资料,不仅可以以文本的格式存储,而且,新型的多媒体介质,如图像、声音和视频,作为原始的分析素材,也日益成为定性分析的新宠。

2.分析方法更加多样:定性方法的种类在最近的二十多年中,更是有了一个质的飞跃。在比较传统的、源自语言学的方法,如内容分析、话语分析、修辞分析、语意分析、符号学、论据分析等方法之外,社会学家也创造出自己独特的定性分析方法,如施特劳斯(Strauss)等人的扎根理论、海斯(Heise)的事件结构分析、拉津(Ragin)的定性对比分析、Abbott 和 Hrycak 采用最优匹配技术的序列分析、亚贝儿(Abell)的形式叙事分析(formal narrative analysis)、鲍尔(Bauer)等人的语库建设、Attride-Stirling 等人的主题网络分析和神经网络技术应用的定性分析领域。所有这些方法的一个共同特征是,把定性研究向更加系统、更加精确、更加严格、更加形式化的方向推进。

3.认识论基础更加多元化:现象学、释义学和本土方法论(ethnomethodology)的认识论,一直是定性分析的大本营,但近年来,实证主义也开始逐渐为定性分析所接纳,解释和阐释之间,由激烈的对立关系,逐渐演变为相互融合。

4.研究过程更加客观规范:定性分析的一个主要问题在于阐释过程中不可避免的主观性。为了尽可能消除"解释者偏见"和主观选择性,定性分析开始遵循严格的程序模板或程序规则,并尝试引入定量分析中的"信度"、"效度"、"代表性"等概念,通过编码和对比,再加上传统的定性分析标准,如可解释性、透明性和一致性,使得定性研究的过程更加规范、阐释的结果更加客观,研究的结论更加可信。

5.研究过程更加有效率:这主要应归功于大量计算机辅助定性数据分析(CAQDA)软件的涌现。从20世纪80年代以来,定性分析过程的数字化和计算机化,已经是一个不可逆转的大趋势。这种发展趋势与定性研究者的理论取向无关,不管他们的理论立场是实证主义、符号互动论,还是本土方法论,大多数定性研究者都在自己的研究当中,开始采用计算机来辅助定性资料的分析过程。据不完全统计,目前已经有二十多种定性分析的软件,分别隶属于德国、英国、法国、美国等国家。其中,有一些软件是国外研究机构的科研成果,可以免费使用,但比较成熟的定性辅助系统大多是商业软件。这些定性分析的辅助系统,不仅使得研究者从处理大量文字材料的繁复劳动中解放出来,而且能够让研究者共享他们各自分析的细节,从而改变了定性研究的流程和研究集体之间的合作方式。同时,由于采用数据库结构,定性资料的管理也更加方便,这就为组织大型定性研究项目(包括多个研究地点、多个研究对象、历时的定性研究)提供了新的可能性。越来越多的定性研究人员开始走出他们的摇椅,坐到计算机屏幕前,湮没在访谈资料和故纸堆中的定性社会学家的形象已经一去不复返了。

6. 定性研究和定量研究的结合更加紧密：在定量分析方法的教材中，定性研究常常被看作是定量研究的前期准备工作，但定性研究者却持完全相反的观点，他们一般认为定性方法是自成一体的，可以完成从形成概念到检验假设的全部研究过程。在实际的应用研究中，定性方法和定量方法常常是交织在一起的，例如，克劳(Currall)等人在研究组织环境重要的群体过程时，通过内容分析把 5 年的参与观察资料量化，然后用统计分析来检验理论假定。格雷(Gray)和邓斯坦(Densten)在研究企业的控制能力时，利用潜变量模型把定性方法和定量方法有机结合在一起。雅各布斯(Jacobs)等人在研究比利时的家庭形态对配偶的家庭劳动分工影响时，首先用定量方法对纵向调查数据进行分析，从定量分析的结果中，又延伸出对核心概念的定性研究。这三个研究分别代表了定量和定性方法相互融合的三个方向：①克劳等人的研究代表着定性方法的实践者试图将定性数据尽可能量化的取向，近年来涌现出的处理调查数据中开放题器的编码问题的工具软件(如 Words at，Smarttext 等，注意：它们都是由著名的统计软件公司出品的处理定性资料的软件)，处理定性资料的传统内容分析软件(如 Nvivo、MaxQDA、Kwalitan 等)也开始提供将定性资料转换到常用统计软件的数据接口，这些工具上的革新将加快这种趋势的发展。②格雷和邓斯坦的工作代表了“方法论多元论”的取向，即在应用研究过程中，通过核心概念的测量模型，把定性研究和定量研究结合在一起。③雅各布斯等人的工作则代表了一部分定量研究者对过度形式化的定量方法的不满，并试图通过定性方法加以弥补。在定量研究领域中，对“模型设定”问题的关注，是定量方法重新试图返回定性研究这种取向的另外一种表现。

与社会调查技术和社会研究方法突飞猛进的现实相比，我国学术界在这些方面的论著的出版似乎显得有些迟缓。虽然已经翻译了美国的一小部分经典定量分析教材，如布莱洛克(Blalock)和巴比(Babie)的教材，也有自己编写的一些教材，如袁方等人的《社会研究原理和方法》、卢淑华的《社会统计学》等，此外，偏重软件操作的还有郭志刚的《社会统计分析方法——spss 软件应用》、郭志刚的《logistic 回归模型——方法与应用》、阮桂海的《spss for windows 高级应用教程》等。在《社会学研究》等专业杂志上，也常常有一些定量分析的应用研究，可是专门的方法和应用模型研究却没有，也没有专门的方法研究期刊。仅就定量研究方法的介绍而言，也存在一些缺陷，主要表现在：

1. 原理和操作脱节。

2. 过分依赖某些商业软件，不全面。

3. 与中国的实证研究相脱节。

4. 不能反映当前方法研究的最新进展。

与定量研究方法相比，由于各种原因，定性研究方法的引进和介绍都比较少。在福特基金会资助的方法高级研讨班上，曾讨论过一些定性研究方法。在定性方法研究方面也有少数专著，如袁方和王汉生 1997 年出版的教程，陈向明 2000 年出版的专著。但总体说来，我们对定性研究方法还停留在初步介绍的阶段，主要的介绍也局限在定性研究的研究设计和资料收集的阶段上，对定性分析方法的介绍则没有能够反映出当代定性方法的最新进展。特别是，在定性分析工具(定性分析软件)的引进和

研究上，基本上还是一个空白。虽然不乏一些出色的定性研究报告，但从方法研究上讲，我们才刚刚起步。当然，我们同时还应当注意到，在历史学领域，我国对定性资料的鉴别、考据和分析，积累了大量的经验和知识，这也应当是定性方法研究的知识来源之一，应努力加以发扬光大。

令人欣慰的是，社会研究方法的引进和出版方面相对滞后的状况终于有所改观。重庆大学出版社的编辑，以独到的学术眼光，逆当前出版界唯利是图的不良选题风气，投入了大量的人力物力，组织出版“万卷方法”。自 2004 年至今，已引进社会科学研究方法方面的专著十余种，在我国社会科学界已经引起了一定的反响。然而，更为可贵的是，重庆大学出版社并未以已经取得的成绩而自满，而是再接再励，在原有“万卷方法”的基础上，进一步组织出版“万卷方法—社会科学研究方法经典译丛”。按我们的设想，“译丛”应该是一个开放的体系，旨在跟踪社会科学研究方法发展的前沿，引进和介绍这一方面的经典著作和最新成果。

“译丛”第一批有《抽样调查设计导论》、《社会科学研究设计原理》、《社会科学研究测量原理》、《社会科学研究分析技术》、《问卷设计手册》、《回归分析法》、《数据再分析法》、《分组比较的统计分析》、《社会网络分析法》、《广义潜变量模型》、《定性变量数据分析》和《复杂调查设计和分析方法》（书名也许有变化）等十余种，几乎囊括了研究设计、测量和分析方法的所有领域，涵盖从基础的回归分析到最前沿的潜变量分析和多水平模型等各种分析方法。无论是社会科学各专业的本科生、研究生，还是社会科学研究的学者都将从中有所收获。

“译丛”由中国社会科学院社会学所社会调查和方法研究室的多位研究人员担纲，主译者都是在社会研究方法各个领域中具有相当造诣的教师和研究人员。“译丛”的译者不仅仅把翻译看作是一个“翻译”，而且也把它看作是一次再学习和再创新。

我们期待“译丛”的出版能对社会研究方法的研究、应用和教学有所推动。

沈崇麟　夏传玲<br>2006 年 12 月于中国社科院社会学所社会调查与方法研究室

# 译者前言

## 一

本书用简洁流畅的笔调为我们提供了有关统计分组比较的丰富信息。它令人信服地表明,统计分组比较方法是社会科学工作者必不可少的工具,灵活运用这种工具可以使研究工作更为深入、更为优美,未来这将是一种具有巨大发展潜力的研究路径。本书并不是对于某一统计模型的入门介绍,因此读者最好对于常用的统计模型具有最基本的知识,但是即使已经对于某一统计模型非常熟悉的读者在阅读本书后仍然会得到新的启示。作者以"分组比较"这一主线将众多统计模型串连在一起,既有一以贯之的思路,又有纷繁多样的变化,许多细节值得人再三回味。将本书与介绍某一统计模型的专门书籍结合起来阅读,读者的收获一定更大。

社会世界是纷繁多样的,在某种程度上多样性反映了一个社会的活力和魅力。在现实世界里,我们享受着这种多样性,我们无法忍受一个单调乏味的世界,我们希望拥有更大的范围可以自由选择,我们用想象力和创造力使世界更为多样化。但是在理论世界中,我们需要的是足够简洁优美的东西。没有人会需要一张 1∶1 的地图,人们在理论解释中真正需要的是过滤掉细节之后的关键信息,它最好能够以尽可能简明的形式表明尽可能多的信息。在研究工作中,我们都会面临这样一种抉择,必须在理论形式的简明性与现实信息的丰富性之间权衡取舍。作者在此表明,恰当地运用统计比较方法,可能是对这一问题的重要解决之道。

统计方法从根本上讲就是比较。在作为统计学基石的假设检验中,我们就是根据数据来比较不同的假设(原假设与备择假设)。在统计学中,一种解释的可信度并非来自于它本身具有的绝对解释力,而是来自于它相对于其他解释所具有的竞争力。研究者需要有很好的判断力来列出需要严肃对待的各种备择假设。只有考虑到其他各种各样复杂精巧的备择假设,我们的研究才能具有最大的说服力。对数据只施以一个模型,只用一种解释来对待,这样得出的结论是武断的。

当然,本书中的统计比较并非指对于不同假设的比较或者对于不同模型的比较,它的核心在于"分组统计比较"。分组是客观存在的:社会中存在着各种自然的和人为的"分组",如不同性别、不同阶层、不同集团、不同国籍等。同时,分组也

是一种认识世界的方法,是我们认识多样性的一种便利之道。尽管组内仍然存在差异,但是通过分组,我们用尽可能少的参数掌握了尽可能多的变异性,进而加深了我们对于现实世界的理解能力。本书涉及的统计模型涵盖了线性模型、广义线性模型、结构方程模型、类别潜变量模型及多水平模型等,此外还涉及一些非参数比较方法。这些模型的复杂程度不同,适用范围不同,但是从分组比较的角度来看,却有相通之处。

统计比较方法应当是具有创造性和富于洞见性。社会统计学的初学者往往最渴望得到的是一种“菜谱式”的教程:什么情况下使用什么模型,然后得到什么结果。计算机的广泛使用使统计学对于社会研究者来说变得“界面友好”,只要按几下鼠标就可得到一长串的数字结果。这在推广了统计学应用的同时,也潜藏了一种滥用的危险。最优秀的定量社会科学研究工作绝不是这样一种程式性的工作。统计方法的创造性应用是无法以一种简单而直接的方法教授的,它只能从那些优秀的定量研究范例中学习。作者在书中提出的例子正是这样的优秀榜样,它们灵活运用了统计工具对于本领域的实质性问题进行了巧妙解答。我们从中再一次看到,优秀的统计运用是一种艺术。值得一提的是,其中数个案例就来自于作者自己的研究,如基于模型对于比率的分解方法、潜类分析中的比较、贝叶斯模型比较等。

网络已经大大地便利了世界范围内信息的传递。本书还有一大特点,它尽可能地利用网络上的资源,包括一些专用的免费软件及教程、共享数据,以及对某些统计量和统计过程的直观演示。这些资源大大方便了读者的学习。当然,这些网址的变动和更新非常快,本书译者尽可能地对于这些网址进行了更新,但如果读者发现某些网址已经失效,也敬请原谅。

本书的叙述风格上也贯穿了比较的原则。例如,在第 2 章中,作者介绍似然比检验、沃尔德检验、拉格朗日乘数检验时,通过比较与图示的方法非常清晰地说明了三者之间的有机联系和差异之处。其他的例子还包括第 3 章中对于线性回归与方差分析的比较,第 10 章中对于固定效应模型与随机效应模型的比较等。

## 二

本书中列举的模型与方法在复杂性上是逐渐加深的,其中包括了一些比较新的研究思路,尤其值得注意,译者在此进行简要介绍。

第 3 章介绍的是线性模型。虽然大部分读者可能熟悉线性模型中的比较,但其中的“回归分解”可能是一种比较新鲜的思路。第 4 章中的再抽样方法,如排列法和自助法,已经被广泛用于研究中进行检验或者估计标准误。相对分布方法是一种比较直观的方法,如果加入协变量影响后,将会更为有趣。第 5 章中的算术标准化是大家熟知的,但基于对数线性模型的标准化(即清洗方法)和分解方法具有更大的灵活性。第 6 章中的广义线性模型提供了一种涵盖众多模型的更具普适性的框架,其中对于考克斯比例风险模型的比较因为涉及基准风险函数,更具复杂

性。第 7 章中的三个问题都是比较前沿的:用于配对数据的条件对数单位回归(conditional logistic regression)可以视为广义线性模型的特例,但在某些软件中的操作需要一些技巧(如在 SAS 中用考克斯回归程序估计条件逻辑斯蒂回归模型);离散异质性的处理需要估计离散系数或者运用双重广义线性模型,读者最好结合相关文献一起阅读;贝叶斯方法引入了模型空间的概念,将原先的单个假设拆分为多个子假设,并考虑到它们的先验概率(或比例权重),这样使得统计分组比较更为细致化。第 8 章中除了对结构方程模型中的协方差结构进行比较外,还引入了均值与协方差结构,这是一种在实际研究工作中非常有用的扩展,因为研究者常常想比较各组中同一潜变量的均值。第 9 章中处理潜类模型中的比较,而对数线性参数化形式的引入使得比较内容比原先经典参数化的比较内容更为丰富。第 10 章中表明"分组"是作为固定效应还是随机效应,这个问题虽然有一些规则可循,但并非绝对;在多水平分析中,如果某一水平中的组数太少,那么将它作为外部变量引入多水平模型中是更为明智的选择。

## 三

最后,译者就几个译法方面的问题向读者作一交待:

1. 英文中有一系列涉及关于"比"的词汇,如 rate, ratio, odds, odds ratio, proportion 等,这些词汇之间在英文中存在细微差别。Rate 指某一事情在一定时间段内的发生数量,如死亡率、发病率等,故而译为"比率"或"率"。Ratio 指两件事物的数量关系,如性别比等,故而译为"比"。Odds 指某一事件的发生机会,如赛马中的赔率,故而译为"几率"。Odds ratio 这一统计学中的重要概念由此译为"几率比"。Proportion 表示部分与整体的关系,按传统译为"比例"。

2. 一些统计模型的名称现在有多种译法,如 logistic, logit, probit 三个彼此有关的词汇就有多种译法。在本书中, logistic 音译为"逻辑斯蒂";logit 在英文中是"logistic probability unit"的缩写,意为"对数概率单位",故而在本书中译为"对数单位";pobit 在英书中是"probability unit"的缩写,意为"概率单位",故而在本书中译为"概率单位"。

3. control 在文中依据上下文有两种译法。文中将 control group 译为"对照组",指没有接受处理的实验对象(或研究对象);而 control 作动词时表示统计学中对混杂因素的处理方式时,译为"控制"①。

4. 不同学科背景对同一方法往往有不同称呼,这种情况下译者只能根据上下文背景取其中一种。如 cohort study 在医药统计学中常译为"队列研究",但在人口

① 关于 control 在统计学中的多种意义,读者可参阅《统计学》(中国统计出版社,Freedman, D. 等著,魏宗舒等译,1997:12-13)。

统计学中常译为“同期群研究”,本书中取前一种译法。再如 case-control study 最早用于医药统计学中,在这一学科背景中通行的译法是“病例-对照研究”,但是现在这种研究设计在其他学科中也日益普及,如果继续沿用“病例-对照研究”这一名称,很容易引发一些歧义。考虑到这一种研究方法对于各个学科都具有普适性,本文将其译为“个案-对照研究”。

5. 书后的索引列出了其中用到的术语及相应译法,请读者参考。

翻译的过程也是译者学习的过程,译文中的不当之处还请读者批评指正。

高　勇

2006 年 10 月

# 作者前言

搞科学研究，不论是自然科学还是社会科学，必不可少地要进行比较。统计方法是比较的核心，因为它们不仅帮助我们理解周围的世界，而且还经常有助于确定如何执行我们的研究。

进行分组比较必要性的最佳示例就是实验，实验可以清晰地定义统计方法。但是，在社会、政治、教育、生物医学等领域中，真正的实验往往是不可能的。使事情更为复杂的是，与结果相关的因素极其纷繁多样。同时，正是这些纷繁多样的分组构成了饶有趣味的比较基础。近年来，特别是在社会、经济、政治科学中，对于多样性的认识和兴趣不断增长，而且这种兴趣在 21 世纪中只会增加不会减少。对于比较方法的兴趣和需求也反映到许多统计方法中，发展这些方法就是为了在进行比较时捕捉到在成分、结构等方面的纷繁多样性的影响。

本书的目的是把近年来发展出来的用于比较的一系列统计方法汇集在一起，对它们做系统的介绍。因此，本书涵盖了由浅入深的一系列主题，从最简单的均值比较到新近发展出来的统计方法，包括贝叶斯方法和层次方法。很自然地，数据分析者最经常想要检验的是一个关于不同结果的假设，本书描述的许多检验反映了此种意愿。但是，研究者对描述与解释差异也很感兴趣；基于此种原因，本书也涵盖了一些分解方法，特别是用于线性的和对数线性的对比率（rate）的比较方法。

本书中的多数例子来自于社会、政治和经济科学，一些例子来自生物医学。但是由于统计方法的广泛性和普适性，许多领域的研究者——不仅是所用例子涉及的社会、政治和生物医学——都会发现本书很有用。研究生和研究者将会发现本书是一本使用方便的参考书，在许多需要进行比较的研究情境中都能用得上。本书也可以作为主要教材在有关统计比较的研究生研讨班上使用，前提是参加者要熟悉线性模型、广义线性模型和潜变量模型的一些形式。在讲授这些专项模型的课程中，本书可以作为有关比较方法的补充读物。

虽然本书涵盖的方法范围广泛，但是它肯定没有囊括所有的比较方法。如果说在选择主题上存在什么偏向的话，那就是更偏向那些新近发展出来的方法。例如，只是有选择性地涉及大家都很熟悉的并与线性模型相关的比较技巧，所包括进来的比较技巧都是最具代表性和最广泛使用的。另一方面，对广义线性模型、潜变量模型、多水平模型则进行了详尽讨论。本书中包括的某些比较方法，如双重广义线性模型和相对分布方法，只是在最近几年才发展起来并加以改进，它们需要使用专业软件。我的希望是把这些方法介绍给更广泛的应用研究者群体，因为我相信它们对于我们的比较研究有着巨大的潜力。

比较方法一直是我的研究兴趣之一。但是直到 1999 年我才想到要撰写本书,当时我在从伊利诺州立大学到剑桥大学的学术假期中。也许正是这次跨越大西洋的旅居使我更为深入地意识到了比较的重要性。但是真正促使我动手的,还是威利出版社的资深编辑斯蒂芬· 奎格利(Stephen Quigley)先生对我的撰稿提纲的热情反应和他的远见卓识。有 8 章是在 2000 年秋季写就的,当时我在伊利诺州立大学教授有关比较的应用统计学研究生课程,我在埃塞克斯大学完成了最后的手稿。

我对为本书写作提供过便利或帮助的人们表达我的谢意:肯· 博伦(Ken Bollen)和诺埃尔· 克雷斯(Noel Cressie)对于撰稿提纲给予了富于建设性的反馈;斯科特· 朗(Scott Long)阅读了 7 章初稿,并给予了极有价值的评论;还要谢谢我的统计比较研讨班上的学生们,他们在我用本书手稿作为教材的班上起到了"试验品(guinea pigs)"的作用,给予我很多反馈,特别要感谢林戈· 穆恩· 霍(Ringo Moon-Ho),他在手稿中找到了种种笔误、不准确之处以及种种错误;感谢汤姆· 斯奈德斯(Tom Snijders)允许我在 10.3 节和 10.4 节使用荷兰学校的数据;感谢吉姆· 克鲁吉尔(Jim Kluegel)允许我在 10.4 节使用国际社会正义的数据;感谢马克· 汉德科克(Mark Handcock)向我介绍相对分布方法;感谢格雷厄姆· 厄普顿(Graham Upton)在书稿校样中找到了一些数学错误。尽管如此,我想本书仍然可能存在一些错误,这些错误都应由我本人负责。

我还要感谢威利出版社的斯蒂芬· 奎格利、希瑟· 哈瑟康(Heather Haselkorn)和罗莎琳· 法卡斯(Rosalyn Farkas),以及约瑟夫· 法恩曼(Joseph Fineman),感谢他们在这一项目中给予我的理解和支持。最后,我的谢意献给我的家庭,庆子和阿西娜,她们的爱和支持使我可以集中精力于这一项目,我愿将本书献给她们。

廖福挺

2002 年 1 月于英国科尔切斯特

目录

第 *1* 章

# 导 论

Introduction

## 1.1 统计比较的基本原理

我十一岁时读过一本书，讲的是一个没有摩擦力的幻想世界，在那里人和物都不能呆在原地不动。现在试想一个没有**比较**的世界。如果我们被剥夺了比较的能力，后果会一样严重。我们会读到手的任何书，会买超市里储存的任何百货，最糟糕的是，如果去买二手车会得到十足的次品！我们不会有任何竞技运动，因为竞争是毫无意义的，而“更快”、“更高”或“更远”并不一定意味着“更好”或“更可取”。

这样一个世界不可想象。事实上，比较是我们生活中不可缺少的部分。正如琼• 希金斯(Joan Higgins)所言，“比较行为是我们日常生活的一部分”，正如父母亲会比较他们的孩子“来看他们是否太胖还是太瘦，太高还是太矮”(Higgins，1981:7)。没有人会否认进行比较是人类思维过程中最重要的活动之一。没有此种活动，世界将成为一个十足无趣的生活场所。比较的过程是平凡的，但在我们的日常生活中又是不可缺少的。

尽管每个人都具有比较区别的能力，但大哲学家们仍然论述了比较的重要性，无非是因为没有比较的存在，就没有人类知识的存在。例如，笛卡尔(Descartes)曾经说过，“只有通过比较，我们方能准确地了解真理。……所有知识，只要不是通过对孤立事物的简单而纯粹的直觉获得，那就必须通过对其中两个或多个事物的比较而获得。人类理智的几乎全部努力无疑就在于使得这一操作变为可能”(Descartes，1963:168；转引自 Lacour，1995)。因此，在笛卡尔看来，知识可以通过两种途径获得：通过对孤立事物进行简单的、纯粹的、直觉式的观察；或者通过比较，此时孤立事物不是知识的对象，不是纯粹而简单的理解，而是理论反思产生了知识。对于笛卡尔来说，我们可以直接地、通过直觉观察的唯一事物就是空间的延展；而对于所有其他事物，包括思维者自身，我们都必须求助于比较方法。在本质上非直觉性的文化、社会，或者别种抽象事物成为我们的知识，并不是通过它们本身而是通过比较的方法。对于

笛卡尔来说,理智的基础就是比较。

比较的领域并不限于文化事物和社会事物。另一位大思想家,歌德(Goethe),毕生都致力于在解剖学、植物学、地质学、矿物学、骨骼学等领域中组建一种"比较形态学"(Lacour,1995)。他在下面一段话中论述了比较对于自然现象研究的核心地位:"自然史是基于比较之上的。外在特征固然重要,但是对于将有机体恰当分离后再重组的任务来说并不充分。"(Goethe,1982:170;转引自 Lacour,1995)歌德是科学比较的实践者。比较原则能够揭示出潜在的连续性,正是这一原则指导着他做出了著名的发现,他识别出了自然史中连接着猿与人的缺失环节。他对于猿骨和人骨的比较研究显示了动物与人在生理结构上有着共同特征,故而有着共同的谱系。歌德认为,任何人只要进行观察和比较都可以做出此类发现,比较是对经验视野的重要补充(Lacour,1995)。歌德的有所发现,是通过对外在可见的特征进行比较,进而揭示出了孤立来看不易察觉的东西。歌德不仅将比较方法应用于相似的事物,而且应用于相异的事物,以便获取知识和发现自然界的潜在原则。

与歌德一样,笛卡尔通过比较方法对于科学知识做出了巨大贡献。为了便于比较,他提出了诸如笛卡尔坐标系和多项式方程这样的分析工具。他使用简洁的符号来代表事物,以便在经验世界中进行比较。笛卡尔的分析方法直到今天仍然是进行统计比较的基础。

本书介绍的是在应用研究中如何进行统计比较。我们比较事物的目的,既为找出其相似之处,也为找出其相异之处。当然,比较从来都不只是发现相似、相同或相等,它也不只是发现差异、不同或不等。有人将比较视为一种辩证过程,在看似相似的事物中寻求相异,在看似相异的事物中寻求相似(Mitchell,1996)。从这种观点看来,给学生的论文作业应是"比较与对比",这样通过"对比"一词就把原本已经隐含在"比较"这一概念中的意思明确化了。然而,我选择沿一个连续体来考虑相似与相异。也就是说,比较行为产生的结果是从完全相同到截然不同的一个序列。我们用这样的方法进行比较的统计检验,检验结果要么保留相等的假设,要么支持不相等的假设,给出的证据力度则各有不同。

此外,本书所讨论的各种方法既适用于横剖面数据也适用于纵贯性数据。也就是说,我们以完全相同的方式来处理同一时点不同组别的差异和不同时点之间的差异。时间产生差异,正如组别变化产生差异一样。这两种差异是否有本质的不同呢?根据鼎鼎大名的美国哲学家约翰·威廉·米勒(John William Miller)的观点(Miller, 1981),答案是明确的"是"。在他看来,历史学和物理学在如何看待时间和差异上存在着根本性的区别。在历史学中,时间的变迁产生了差异,所有这种差异都是原生的,故而是本体性的。在物理学中,目标是解释一事怎样地不同于另一事,故而差异是衍生性的而非本体性的。对于我们进行统计比较的目的来说,时间变迁产生的差异和同一时点上事物间的差异都是衍生性的;但是这种衍生性有助于对于社会和自然的基本

机制的新颖理解。新知识则是本体性的，至少有望如此。

## 1.2　社会科学中的比较研究

在社会科学中，一种正在日益盛行的做法是比较社会群体间的人类行为。其关注的行为可以是经济的、政治的、心理的或社会的，关注的层面可以是个人的也可以是组织的，关注的群体往往有种族的、民族的、性别的或国家的。近年来，理解和强调多元文化主义、包容性、全球化的社会趋势在 21 世纪将进一步促进比较研究的发展。虽然下文中征引的例子局限于社会科学中，但统计比较在其他领域中已被广泛使用，特别应用于生物科学和医药科学。下面的讨论只展现出了某一座冰山之一角，更不用说还有众多的冰山存在。比较研究的普适性和重要性即便如何高估也并不为过。

比较研究的起源可以追溯至社会科学发端之时。许多经典社会科学著述者研究和比较了当时可据可征的所有社会形态，对于他们来讲，“比较”研究的概念颇显多余，因为社会科学从定义上讲就必定是比较的(Nowak，1989)。韦伯(Weber)的社会理论为跨国比较提供了坚实基础，这一点事实上已经得到广泛公认(Wesolowski，1989)。今天对于比较研究的强调，只不过代表了对于社会比较的兴趣的一种复兴，这一点在下面给出的参考文献中可以得到证实。

对于众多领域的社会科学家来说，进行实质性的信息的统计比较已经成为一种重要的分析策略。人类学、经济学、心理学、政治科学、社会学中的研究者一直以来就对人类群体、制度、组织和社会之间进行系统比较颇感兴趣。从经验研究的角度来看，大多数比较可划分为两大类——性别、民族和种族群体之间的比较，跨文化和跨国家的比较。

无论是哪一种的例子在社会科学文献中都比比皆是。许多新近的研究关注经济、政治、心理和社会行为上的性别差异(如，Arber and Cooper，1999；Echevarria and Merlo，1999；Henriques and Calhoun，1999；Johnson and Marini，1998；McDowell，Singell，and Zilliak，1999；Perrucci，Perrucci，and Targ，1997；Sapiro and Conover，1997)。在同样的社会科学领域中，我们也可以发现对于一系列行为上的种族差异和民族差异的研究(如，Choi，1997；Henriques and Calhoun，1999；Lloyd and South，1996；Mennen，1994；Telles and Lim，1998)。以上参考文献只不过反映了考察种族、民族和性别差异的社会科学经验研究这一巨大冰山之一角。

跨国家、跨文化研究的趋势在近十年中取得了更大的发展势头。属于这一传统的许多研究不可小视，已出现在社会科学的顶级期刊上(如，Alderson and Nielsen，1999；Bornstein，Haynes，and Azuma，1998；Browne，1998；Ehrhardt-Martinez，1998；Ellina and Moore，1990；Hupkens，Knibbe，and Otter-

loo,1998; Ishida, Muller, and Ridge,1995; Kenworthy,1999; Sullivan and Transue,1999; Westholm and Niemi,1992; Wright, Baxter, and Birkelund,1995)。在经济、社会和行为科学中跨国研究的例子数不胜数,而比较多个社会的教育研究在文献中也已经频繁出现(如,Becker,1992; Cai,1995; Silver, Leung, and Cai,1995)。

使用多国数据进行的社会经验研究可以根据研究者的意图划分为以下四种:国家作为研究对象,国家作为研究背景,国家作为分析单位,国家处于跨国场景中(Kohn,1989)。在社会学中虽然具有全球化信念的研究者们对于跨国研究的有效性提出了批评(如 McMichael,1990; Martin and Beittel,1998),但在更广泛的社会科学中,跨国研究或比较研究正在越来越引起人们的普遍关注,前面所提及的新近的例子可作为佐证。在单个社会科学领域中给予比较研究的特殊关注也证明了这一点,例如,几年前在《现代人类学》(Current Anthropology)上的讨论;再如,由阿茨(Arts)和哈尔曼(Halman)编辑的新书讨论了定量比较社会学的新近方向(Arts and Halman, 1999)。

但是,还没有人系统地论述过如何用定量方法进行有效的分组比较。社会科学研究者在进行社会比较时往往忽略任何正规的检验(上面列举的许多文献也是如此)。一种解释可能是:当研究者学习某一种特定统计模型时,通常进行比较并不是关注重点;此外还有一个原因是大多数应用统计学课本中几乎没有涵盖这一主题(除了比较不同组的均值和比例以外),更不用说以一种系统的方式加以讨论了。

本书关心的是组间的统计比较。大多数时候我们关注的焦点是参数方法和参数模型,虽然也会包括一些非参数方法。参数方法的主要优点是:同一检验可应用于多个不同的模型,只要参数方法的前提假定仍然有效。此外,非参数检验在较简单的分析中发展程度更高,而在较高级的统计模型中发展程度不高。同样,我们主要将关注统计学中的频率论方法,因为就比较的目的而言它的发展程度更高;而贝叶斯方法只是作为一种扩展而加以介绍。

## 1.3 本书关注的重点

比较研究在许多领域中都是一种重要的研究方法。由沙伦·安德森(Sharon Anderson)及其同事在威利出版社出版的一本颇有影响的书(Anderson et al, 1980),总结了许多用于比较目的的统计方法。其中,考察的方法包括标准化、配对、方差分析、对数单位分析、对数线性分析、生存分析。这本书在许多年来帮助了研究者在处理混杂变量问题时进行比较研究。

但是准确地讲,正如其标题"比较研究的统计方法:偏误消减技术(Statistical Methods for Comparative Stadies: Techniques for Bias Reduction)"所指出的,那本书所关注的重点是比较中的种种混杂因素(confounding factors)。相比之下,本书假定

研究者对于潜在的混杂因素已是了然于胸,并且已把它们都纳入了所研究的统计模型中,不管它是线性的、广义线性的还是广义线性混合模型。唯一的例外是对比率的比较,通常该方法并不将参数所捕捉到的混杂因素的影响纳入模型之中。因此,我们将对比率的比较问题单独进行考察。

的确,统计学的发展和它在众多生物医学和社会科学中的应用已经使得混杂变量的潜在后果和通常处理办法变得非常普及。某些混杂因素导致的偏误可以用模型中的参数来表达,它们在不同组之间的重要程度同样很有趣。这些参数,连同代表其他核心变量影响的其他参数,构成了要进行比较的参数向量。因此,我们在本书中关注的重点将是比较方法本身。

## 1.4　本书的概要

在导论之后,本书在随后的 9 章实质内容中探讨了组间的统计比较。虽然在多水平分析中我们会讨论随机分组,但在许多探讨中,分组被认为是固定的。在社会和生物科学中,将分组视为固定的,这在绝大多数情况下是一种明智的方法,也是通常所实行的处理方式。在第 10 章中,我们将讨论从随机分布中得到的分组。

### 1.4.1　第 2 章:比较的统计学基础

本章介绍了在许多统计模型中进行统计分组比较的系统方法。并考察了六种常见的检验统计量:t,$\chi^2$,F,似然比,沃尔德,拉格朗日乘数。后三种是渐进等效的。

### 1.4.2　第 3 章:线性回归中的比较

本章首先介绍了最简单的比较:分组的均值比较和比例比较。然后,再把它们看做线性回归中检验参数相等性的一个特例。我们还探讨了一些多重比较技术:方差分析、协方差分析和回归分解。

### 1.4.3　第 4 章:非参数比较

本章我们探讨了几种主要的非参数比较方法。其中包括非参数检验和再抽样方法(该方法也可应用于其他参数检验)。同时还讨论了相对分布方法的最新发展。

### 1.4.4　第 5 章:比率的比较

社会研究者和生物医学研究者常用比率进行比较,同时也用它来评估事件发生的密集度。成分结构常常会混淆比较结果。在本章中,我们将介绍标准化方法和分解方法,每一种方法又可分为传统的算术类型和基于模型的

方法。

### 1.4.5 第6章:广义线性模型中的比较

本章首先介绍了被称为广义线性模型的模型族;然后又考察了在此类模型中以组为基础的参数比较,诸如对数单位模型、概率单位模型、多项式对数单位模型、风险率模型都是广义线性模型的成员。由于广义线性模型提供了一种将原本不相关的模型联系起来的框架,因此,本章相当重要。

### 1.4.6 第7章:广义线性模型中进行比较的其他问题

我们探讨的三种扩展是:配对研究、离散异质性和贝叶斯模型。广义线性模型的贝叶斯表述可以视为对这些模型的一种推广,异质性可以视为广义线性模型中的一种非标准条件,而配对研究则是广义线性模型用于在某些特征上将处理组和对照组配对的设计中时的特例。此类设计不仅在进行实验的学科中有用,而且在实施"类实验"的学科中也有用。因此,作为统计比较的一种特殊类型,配对研究值得我们特别关注。

广义线性模型中的分组比较有时可能会受到非标准条件的影响,尤其是离散异质性。线性回归中的异方差性只是一种特例,离散的异质性在其他广义线性模型中都可能出现。本章我们将研究一个存在离散异质性的对数单位模型,用"类似然比检验"来进行组间比较。接下来我们介绍贝叶斯统计学的表述,即"广义线性贝叶斯模型",以此作为广义线性模型的一种扩展。在广义线性贝叶斯模型中进行比较时,我们比较的不仅是"各组相同"与"各组相异"这两个相互竞争的模型;与此不同,我们在检验均等假设时还要考虑到模型空间中处于两个极端之间的所有可能的竞争性模型。

### 1.4.7 第8章:结构方程模型中的比较

至此考察的模型都是建立在单个方程基础上的。本章我们探讨多方程系统,其中的内生变量和外生变量既可以是可观察的也可以是潜在的。在此类系统中分组比较仍然是可取的。结构方程建模中的分组比较可以采用不同的形式。就其所探讨的模型而言,假定的参数可能不同;此外,模型假定的结构可能不同;或者两种情况可能同时为真。除此之外,比较可能涉及协方差结构,均值结构,或者同时涉及两者。与用于比较广义线性模型的计算机程序相比,比较结构方程模型的句法要复杂得多。因此,在附录中提供了针对例子而列出的 EQS 和 LISREL 程序。

### 1.4.8 第9章:类别潜变量的比较

在结构方程模型中,潜变量被假定是连续的。但是潜变量同样可以作为离散的来进行概念化和处理。本章将探讨各种类型的类别潜变量模型(models of categorical latent variables),但是将集中讨论在潜类分析(latent class

analysis)中的分组比较,以及将潜(类)变量引入到传统路径分析模型后的扩展中的分组比较。这种潜类模型也可以参数化为对数线性模型,通过这样一种参数化,可以将对数线性模型的所有主要进展纳入潜类分析中。我们将讨论在潜类模型的这些变体中的组间比较。因为有许多种专用软件包可用来分析此类模型,附录中列出了主要的几种软件。另一个附录给出了针对示例而列出的计算机程序。

### 1.4.9　第 10 章:多水平分析中的比较

近年来,多水平分析在医学、教育学、社会科学研究中日趋普及。我们如何在这种分析中处理分组比较? 本章中我们考虑将分组视为一种分析水平,并且将这种处理方法与先前讨论的在广义线性模型中检验相等性假设的处理方法进行对比。在这种方法中,分组可以被视为来源于随机分布,而非固定的单位。

### 1.4.10　总结

总的来说,这些章节中讨论的方法在复杂性上是逐渐加深的。若熟悉某一种方法的研究者对于了解如何使用这种方法进行比较感兴趣,完全可以直接阅读相关章节而不会有任何麻烦。例如,熟悉结构方程模型的研究者可以直接阅读第 8 章,当然,如果能先快速浏览一下第 2 章中的基本知识,阅读效果可能会更好一些。

第2章

# 比较的统计学基础

Statistical Foundation for Comparison

## 2.1 一个统计比较的系统

为了便于采用各种各样的方法和模型进行统计比较,我们定义以下简单而又通用的检验相等与不等的系统。这一系统适用于从最基本(如均值比较)的到最高级(如潜变量模型)的各种统计分组比较。

让我们首先来介绍一下典型的原假设:

$$H_0: S_1 = S_2 = \cdots = S_G$$

统计量 S 是由感兴趣的统计量或者参数组成的向量,它可以是诸如均值、比例、回归系数等。下标 G 表示要比较的组的总数。此假设的检验统计量是形式如下的函数:

$$d(S_g) \qquad 其中\ g = 1, 2, \cdots, G$$

根据分布形态的不同,函数 d(·)分别给出 $t, F, \chi^2$ 检验统计量,它们对应于不同自由度有不同的临界值(cutoff values)。不论函数 d(·)的形式如何,统计比较的目的都是为了显示零假设是应该加以保留还是拒斥。绝大多数情况下,这一函数都服从某一参数分布,但是它也可定义为代表非参数检验或者其他。例如,函数也可以是排列检验函数。

依照各种参数模型服从的通则,本章我们首先关注的是 $t, F, \chi^2$ 检验统计量。读者将会明白,许多参数方法中的统计分组比较服从相同的基本原理,使用相同的统计检验。非参数检验和排列检验问题将在随后的章节中适时讨论。

## 2.2 检验统计量

以下提到的检验都是所谓的“显著性检验”。在此类检验中,假定数据来自随机抽样。所有的显著性检验都一样,如果你拥有整个总体,那么任何差异

都是真实而又显著的。如果假定数据来自非随机抽样，那么统计显著性将无法确定，虽然有时显著性检验被应用于得到一个对于关系强度（或对差异程度）粗略的"经验法则"。值得注意的是，下面描述的检验均基于各自的分布。但这些分布本身并不构成检验，而只是描述其检验行为。对于每一种检验，我们首先介绍它的分布情况，然后再介绍其检验。

### 2.2.1　$\chi^2$ 检验

假定变量 Y 服从正态分布，我们拥有此变量 Y 的个案总体。然后，从这一总体中抽取样本，每次抽一个个案，N = 1。对于每一个样本，通过计算它的 Z 值并将其标准化，然后取其平方，得到 $Z^2$。我们称平方后的标准值为 $\chi^2_{(1)}$，即：

$$\chi^2_{(1)} = Z^2 \tag{2.1}$$

这样我们就从服从标准正态分布的总体中得到了一个 $\chi^2$ 分布。式(2.1)给出了一个自由度为 1 的卡方分布，众所周知，其分布曲线呈向右陡降的形状。由于 $\chi^2$ 是一个平方量，其取值必为非负实数，范围从零到正无穷。

现在假定从先前的总体中独立而随机地抽取两个个案。通过如此重复抽样，每次都计算两个标准化取值的平方之和，则可得随机变量 $\chi^2_{(2)}$：

$$\chi^2_{(2)} = Z_1^2 + Z_2^2 \tag{2.2}$$

即自由度为 2 的卡方分布。$\chi^2_{(2)}$ 的分布比 $\chi^2_{(1)}$ 的分布的偏斜度要略小一些。

我们进一步假定从原先描述的正态分布中抽取 N 个独立随机观察，由此可得随机变量 $\chi^2_{(N)}$：

$$\chi^2_{(N)} = \sum_{1} Z_i^2 \qquad i = 1, 2, \cdots, N \tag{2.3}$$

$\chi^2_{(N)}$ 的分布形状取决于每次随机抽取的独立个案的数目。从一个正态分布总体中每次随机抽取 N 个独立个案，这些个案的标准化取值的平方和则服从自由度为 N 的卡方分布。

由于 $Z_i = (y_i - \mu)/\sigma$，因此

$$\sum_{1=1}^{N} \left(\frac{y_i - \mu}{\sigma}\right)^2 = \frac{\sum_{1=1}^{N} (y_i - \mu)^2}{\sigma^2} \sim \chi^2_{(N)}$$

我们可以用矩阵符号将上述结果概括地表达为：

$$\frac{1}{\sigma^2}(y - \mu)'(y - \mu) = (y - \mu)'(\sigma^2 I)^{-1}(y - \mu) \sim \chi^2(N) \tag{2.4}$$

中间的因子 $\sigma^2 I$ 是 $y_i$ 的协方差矩阵，因此，可以用更一般性的 $\sigma^2 W$ 来替代。这对于某些检验来说将是非常有用的形式，如在随后将叙述的沃尔德检验中。

与后面要讨论的 t 分布一样，卡方分布不是一种单一分布，而是对应于某个特定自由度的分布。当自由度较小时，分布曲线严重向右偏斜；当自由度较大时，分布曲线或多或少地趋于对称。这种现象可用正态分布中大多数个案

所处的位置来理解。例如,在标准正态分布中有约68%的个案位于 -1 ~1,这意味着相对应的 $\chi^2_{(1)}$ 的取值为 0 ~1,由此便导致了非常偏斜的分布形状。由于要对两个或多个独立抽取出来的随机观察进行加总,相对应的 $\chi^2_{(2)}$ 或者更一般性的 $\chi^2_{(N)}$ 中就有大量个案从零转移为更大的取值,故而显示出更为对称的分布。在互联网中位于 http://faculty. vassar. edu/lowry/csqsamp. html[①] 上的卡方抽样分布生成器是一个直观的工具,不仅有助于理解分布形状与自由度之间的关系,而且有助于直观理解与此相关的各种0.05 ~0.001 常见显著度下的临界值。

有许多显著性检验都依赖于卡方分布。最著名的可能是皮尔逊卡方检验,它被广泛地应用于列联表分析中。随后将要讨论的似然比、沃尔德、拉格朗日乘数检验统计量,也都服从卡方分布。

### 2.2.2 t检验

令 $\chi^2_{(k)}$ 为自由度(df)为 k 的卡方变量。此变量除 k 即可得到修正 $\chi^2_{(k)}$ 值(Wonnacott and Wonnacott 1979):

$$C_k^2 = \frac{\chi^2_{(k)}}{k}$$

$\chi^2_{(k)}$ 可以被视为一种样本和,因此 $C_k^2$ 便可被视为从总体 $Z_k$ 中得到的样本均值,期望值为1,方差为2/k。自由度为 k 的 t分布的数学定义为:

$$t_k \equiv \frac{Z}{C_k} \tag{2.5}$$

t分布也被称为"学生分布",这是为了纪念 W· G· 戈瑟特(W. G. Gossett),他在20世纪早期以"学生"为笔名首次公布了这一分布。那时,他的雇主规定他不得以真实姓名发表其研究成果。

t分布与正态分布具有某些共同的特征。两种分布都是对称的,范围都从负无穷延展到正无穷。但是 t分布与正态分布的不同之处在于:它的形状要依赖于自由度的数目。与此不同,Z 比值是不依赖于样本量或自由度的,服从标准正态曲线。由于自由度数目的范围是从1到无穷,故 t分布的形状则与正态分布不同。自由度为1的 t分布与标准正态分布的偏离最大,而当自由度的数目增加时,学生分布越来越接近标准正态分布的形状。当自由度 df =30 时,它与标准正态分布就更不易区分了。当自由度 df = ∞ 时,两种分布是完全相同的。在这种意义上,我们可以将正态分布视为 df = ∞ 时学生分布的一种特例。

t统计量与 Z 统计量密切相关,Z 统计量是离差统计量(其自身也是一个

---

① 原文提供的网址为:http://Vassun. Vassar. edu/ ~ lowry/csqsamp. html,似乎已变更。——译者注

随机变量）与抽样分布的标准误之比。Z 比值服从标准正态分布。Z 比值假定总体方差已知，故统计量的标准误也已知。但是在多数情形下，我们只有样本数据，而根本不知道总体方差的真值。因此，我们用统计量的样本方差来估计其标准误。一个统计量与其估计的标准误之比构成了 t 统计量。t 检验是基于 t 分布基础上的，t 分布是 t 比值的期望分布。

通常 t 统计量被用于检验两个组间均值之间的差异，对于组 g 和 h，形式如下：

$$t_{df} = \frac{(\mu_g - \mu_h) - 0}{\hat{\sigma}_{\mu_g - \mu_h}} = \frac{\mu_g - \mu_h}{\hat{\sigma}_{\mu_g - \mu_h}} \tag{2.6}$$

在此我们假定原假设认为差异为零，故从分子中省略了零差异。

当我们用样本方差 $\hat{\sigma}$ 替代了 $\sigma$ 时，就得到一个“拟标准化”变量（quasistandardized variable）。一般而言，只要任何正态变量被拟标准化了，它就变成一个 t 变量而不是一个 Z 变量。回归系数 $\hat{\beta}$ 就是一个极佳的例子。拟标准化变量是一个随机变量，其生成的置信区间也将比由 Z 生成的置信区间要宽一些。

## 2.2.3　F 检验

F 分布，或者随机变量 F 的分布，是因罗纳德· 费希尔爵士（Sir Ronald Fisher）而得名，他创造了这一概念以及这一分布的一些主要应用。自由度为 $df_1$ 和 $df_2$ 的 F 变量的数学定义与 t 变量的定义类似：

$$F_{[df_1, df_2]} \equiv \frac{C^2_{df_1}}{C^2_{df_2}} \tag{2.7}$$

其中，$C^2_{df_1}$ 和 $C^2_{df_2}$ 是彼此独立的修正卡方变量。当它们完全相关时，F 比值为 1，则不再是一个随机变量。

F 的另一个常见用途是用于检验等方差性。通常我们需要比较两个总体方差，F 分布就是它们估计方差之比的分布。试想有两个不同的总体，每一个都服从变量 Y 的正态分布。这两个总体的均值不同而方差相同，都是 $\sigma^2$，我们可从这两个总体中抽取两个独立样本。这两个随机样本的样本量分别为 $N_1$，$N_2$。对于从两个总体中得到的每一对样本，我们计算估计方差 $\hat{\sigma}^2$，或者 $S_1^2$ 比 $S_2^2$，它服从 F 分布：

$$F = \frac{S_1^2}{S_2^2} \tag{2.8}$$

注意，由于它们的总体方差被假定是相同的，因此通过计算两个方差估计值之比，实际上得到的是两个独立的卡方变量除它们各自的自由度之比。即：

$$F_{[df_1, df_2]} \equiv \frac{\chi^2_{(df_1)}/df_1}{\chi^2_{(df_2)}/df_2} \tag{2.9}$$

这样一个随机变量被称为 F 比，服从 F 分布。姑且不管 F 比的计算在数学上有多么复杂，只要记住 F 的密度取决于两个参数 $df_1$ 和 $df_2$，即比值的分子

和分母为各自的自由度数目。分布的形状是两个自由度数目的函数。为了理解 F 密度函数和自由度之间的关系,网页 http://faculty.vassar.edu/lowry/fsamp.html[①] 会很有帮助。与卡方抽样分布生成器一样,使用者只要提供自由度就可以从中得到一个相应的 F 密度函数。

F 检验应用广泛,特别是在方差分析和线性回归中。在有 k 个参数的线性多元回归中,检验参数同时相等假设的 F 统计量形式如下:

$$F_{[df_1, df_2]} = \frac{SSM/(k-1)}{SSE/(N-k)}$$

即模型的平方和与误差项的平方和之比,分别除以它们各自的自由度,则构成 F 比。

### 2.2.4 似然比检验

接下来我们探讨似然比检验。给定一个包括观察数据$\{n_i, i=1,2,\cdots,N\}$的样本,似然函数 L 是观察到$\{n_i\}$的联合概率对于某些未知参数 θ 的函数。L 由研究者用以分析数据的模型定义,依据模型背后的分布而有某一特定形式。最大似然估计法(以下简称 ML)就是获得使似然函数 L 最大化的估计量,以此作为参数取值的程序。假定我们的原假设为 $\theta_0=0$,原假设可以视为一种对于全体参数施加的限定(restriction)。更为一般化地,我们将参数向量 θ 表示为 S,表示任何要检验的统计量,原假设为 $s_1=s_0$,备择假设为 $s_1 \neq s_0$。现在令 $L_1$ 为没有额外限定时数据的最大似然值。换言之,$L_1$ 为 θ 不受限定,故以 ML 估计值(以 $\theta_1$ 表示)替代时的数据为似然值。令 $L_0$ 为在基于原假设受到限定的向量 $\theta_0$ 情形下的最大似然值。$\theta_0$ 的维数为 $k \times 1$(即 $L_0$ 比 $L_1$ 少 k 个参数)。

接下来,我们可得到比值 $L_0/L_1$。这个比值的取值总是 0~1,原假设为真的可能性越小,比值就越小。这一比值是似然比检验(以下简称 LRT)统计量的基础。在实践中,我们用 $-2\ln(L_0/L_1)$ 来计算在某一置信度下的 LRT。换言之,我们将似然比的自然对数乘 −2。或可通过计算 −2 倍的对数似然值之差,即 $-2\ln L_1$ 与 $-2\ln L_0$ 之差,则 LRT 统计量为

$$LR = -2(\ln L_1 - \ln L_0) \tag{2.10}$$

因为 LRT 统计量服从卡方分布,我们可以通过比较 LRT 统计量与自由度为 k 的卡方分布的 $100 \times (1-\alpha)$ 上侧的百分点(percentile point)来查看检验的显著性水平。通常 LRT 统计量的分布与卡方分布的近似程度较好,即使对小样本来说也是如此。如果依上述方法计算出来的 LRT 的值大于自由度为 k 的卡方分布在百分点处的临界值,那么就拒斥原假设。百分点是分析者根据检验的置信水平来选择的。

---

① 原文提供的网址为:http://vassun.vassar.edu./~lowry/fsamp.html,似乎已变更。——译者注

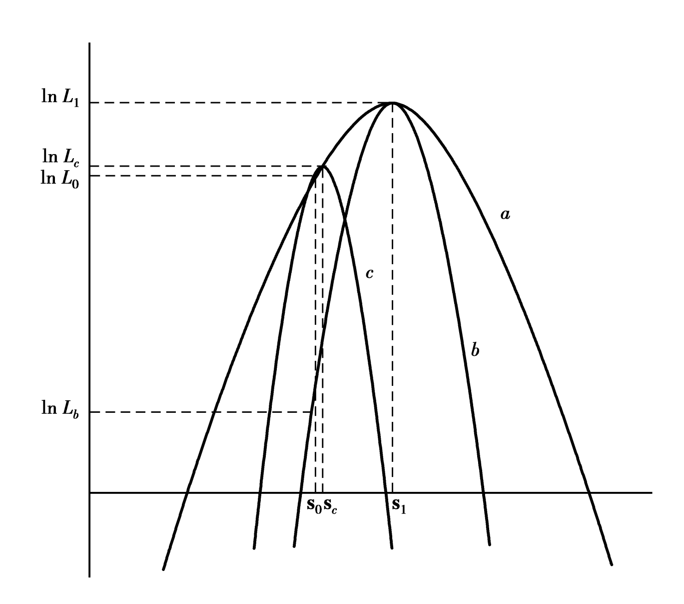



**图 2.1　三种渐进等效的检验方法的比较**

让我们回到一般化的情形中,在 S 中只有一个元素时来评估 S,并且如图 2.1 所示来描述 LRT。假定我们关注的某一特定统计量 S 有 ML 估计 $s_1$,曲线 a 代表统计量的对数似然函数。现在如果 S 的受限定值为 $s_0$,那么 LRT 统计量由 $\ln L_1$ 与 $\ln L_0$ 之间的距离给出,它服从卡方分布。$\ln L_1$ 与 $\ln L_0$ 之间的距离恰是似然比的一半。对数似然函数 a 的曲率可表示为 $C(\theta_1)$,其定义为当 $\theta = \theta_1$ 时 $d^2 \ln L / d\theta^2$ 的绝对值(Buse, 1982)。在 $s_1$ 与 $s_0$ 之间的距离既定时,$C(\theta_1)$ 的曲率越大,则 $\ln L_1$ 与 $\ln L_0$ 之间的距离(似然比的一半)越大;反之,在曲率$C(\theta_1)$既定时,$s_1$ 与 $s_0$ 之间的距离越大,$\ln L_0$ 距离最大值 $\ln L_1$ 越远。这种图示法将同样有助于我们理解沃尔德检验。

## 2.2.5　沃尔德检验

与由 $\ln L_0$ 与 $\ln L_1$ 的最大值之间的距离决定的 LRT 不同,由沃尔德(Wald, 1943)提出的沃尔德检验(The Wald Test)考虑的是 $s_1$ 与 $s_0$ 之间的距离的平方。距离的平方越大,拒斥原假设的证据就越强有力。由前一节讨论可知,曲率会影响 LRT,曲率同样也会影响这种检验过程,因此 $s_1$ 与 $s_0$ 之间的距离的平方必须用曲率 $C(\theta_1)$ 来加权。两组数据可能产生两条具有不同曲率的曲线,但会产生完全相同的 $s_1 - s_0$;如果我们考察 LRT,一组数据会比另一组数据更倾向于支持原假设,图 2.1 有助于说明这一点。曲线 a 与曲线 b 产生于两组不同的数据,但其$(s_1 - s_0)^2$ 的值完全相同。曲线 a 给出的 $\ln L_0$ 的值比曲线 b 更大(曲线 b 中的 $\ln L_0$ 在图中用 $\ln L_b$ 表示)。因此,沃尔德统计量是一个加权后的距离的平方:

$$W = (s_1 - s_0)^2 C(s_1 - s_0)$$

它渐进地服从自由度为 1 的卡方分布。上述公式可用更为一般的形式来代替：

$$W = (s_1 - s_0)^2 I(s)(s_1 - s_0) \tag{2.11}$$

其中

$$I(s) = \frac{d^2 \ln L}{ds^2}$$

$I(s)$ 就是所谓的“信息矩阵”。现在，距离的平方用平均曲率进行了加权。

作为一种在假设检验中广泛使用的对 LRT 的替代方法，沃尔德检验（以下简称 WT）也提供了一种对 t 检验的更为直接的类比，至少在单个参数的情况下如此。对于 $H_0: \theta = 0$，检验具有以下一般形式：

$$W = \hat{\theta}'_1 [\text{cov}(\hat{\theta}_1)]^{-1} \hat{\theta}_1 \tag{2.12}$$

其中，$\hat{\theta}_1$ 代表在非限定模型中的 $\theta$ 的估计值。在 $\theta$ 只有一个元素时（即自由度为 1 时），$W$ 就只是参数估计值与它的估计标准误差之比的平方，在主要的统计软件包的逻辑斯蒂回归程序中都报告这一数值。

沃尔德统计量与似然比统计量是渐进等效的，同时也近似于卡方分布。沃尔德原则基于如下理念：如果对于参数的限定是真实的，那么非限定模型应当近似地满足这一限定。即使计算机输出中没有提供，沃尔德统计量的计算也是相当简单的，它显示了非限定模型与限定理想的偏离程度。

## 2.2.6 拉格朗日乘数检验

拉格朗日乘数检验（The Lagrange Multiplier Test，以下简称 LMT），也被称为得分统计量（the score test），是由 R• A• 费希尔（R. A. Fisher）和 C• R• 劳（C. R. Rao）提出的，与 WT 一样，它是用于检验如 $\theta = \theta_0$ 这样的假设的另一种对 LRT 的替代方法。对于对数似然函数曲率的讨论同样适用于 LMT，不过这种检验要用到的是施加限定时的对数似然函数的性质。如果原假设为真，受限定的 $\ln L_0$ 将接近于不受限定的 $\ln L_1$。由于不受限定的估计值使得对数似然最大化，因此，它们满足条件 $d(s_1) = 0$，其中，$d(s_1) = d\ln L/ds_1$。我们用平方得分 $[d(s_0)]^2$ 作为检验统计量，因为函数斜率的符号无关紧要。

我们看到，两个不同的数据集可以产生有着相同斜率的两个不同对数似然函数，其中一条曲线中的最大值 $s_1$ 更接近于 $s_0$，如图 2.1 所示。假定另外一组数据产生了曲线 c 代表的对数似然函数。新的 $s_1$ 对应于曲线 c 的最大值在 x 轴上的位置，用 $s_c$ 代表。差 $s_c - s_0$ 和 $L_c$ 与 $L_0$ 之差都更小了。曲线 a 和曲线 c 代表的对数似函数在限定值 $s_0$ 处（即两条曲线相切之处）的斜率完全相同。这一现象是由于似然函数的曲率不同导致的。为了纠正这一问题，我们再次用曲率对斜率的平方进行加权。在此情形下，曲率越大，$s_0$ 越接近对数似然的最大值。这表明应当用倒数 $C(s_0)^{-1}$ 来加权，以确保 $s_0$ 与 $s_c$ 之间的距离越近

相应的检验统计量取值也越小：

$$LM = [d(s_0)]^2 C(s_0)^{-1}$$

与前面一样，上述公式可以替换成更为一般化的形式：

$$LM = [d(s_0)]^2 I(s_0)^{-1} \tag{2.13}$$

它渐进地服从自由度为 1 的卡方分布。

更为一般地，检验可被视为基于 ln L 对 θ 的偏导数向量的二次型函数，在假定 $H_0$ 成立的$\hat{\theta}$点进行估值。其数学表达式如下：

$$LM = [d(\hat{\theta}_0)]'[I(\hat{\theta}_0)^{-1}][d(\hat{\theta}_0)] \tag{2.14}$$

其中，$d(\theta_0)$ 包含了对数似然函数在限定值处的导数，即$\partial \ln L(\hat{\theta}_0)/\partial \hat{\theta}_0$；而 $I(\hat{\theta}_0)$ 是信息矩阵。LMT 与 LRT 和 WT 是渐进等效的，虽然它只在限定值处进行评估。

## 2.2.7　对 LRT，MT，LMT 的总比较

LRT 在实践中的一个缺点是它要求对限定模型和非限定模型都进行估计，虽然对于今天的高速计算机来说问题已经不像以前那样严重了。在较为复杂的模型中，这些估计值中的某一个或另一个可能很难计算。如果这一问题阻碍了假设检验，那会十分令人遗憾。幸运的是，WT 和 LMT 规避了这一问题。在检验参数显著性的典型过程中，WT 只依赖于非限定的估计量，而 LMT 只依赖于限定的估计量。这两种替代方法中都基于一个渐进正态分布的估计量，都基于满秩二次型的分布。这三种方法是渐进等效的，但是它们的小样本性质大部分仍是未知的，在小样本中的表现可能差异很大（Greene，1990）。对于大样本而言，它们都接近于卡方分布。

这三种渐进等效的检验是建立在与似然函数相关的非常不同的理论原则之上。回顾一下图 2.1 将显示出 LRT，WT，LMT 之间的不同之处。为了检验 $s_1$ 与 $s_0$，LRT 统计量由 ln $L_1$ 与 ln $L_0$ 之差给出，它服从卡方分布。请注意，对数似然函数被描述为一种估计中不会被多个最大值（multiple modes）或其他此类问题干扰的理想状态。WT 关注的是 $s_1$ 与 $s_0$ 之差的平方，同时由于曲线 a 与 b 之间曲率可能存在差异而加以修正。这两个似然函数由两个不同的数据集产生，曲率的不同可以用对数似然函数的二阶导数的负值大小来测量。在某种意义上，沃尔德统计量是在用对数似然函数的曲率量度来标准化 $s_1$ 与 $s_0$ 之间的距离。相比之下，LMT 测量的则是得分 $d(s_0)$ 的平方。同样，曲线 a 与曲线 c 可能有着相同的斜率，但曲率不同。因此，我们必须通过 $C(s_0)$ 的倒数来调整 $[d(s_0)]^2$，以便抓住曲率与检验统计量取值之间的反向关系。LMT 这样就给了我们另一种进行假设检验的选择。

## 2.3 比较什么

通常数据分析者欲比较各组的参数,就必须通过假定检验参数相等来实现。虽然它肯定是最常见的一类比较,但正如下面章节中要说明的,它只代表了其中一类。此外,研究者感兴趣的方面还包括数据背后的统计分布、基本数据结构、模型结构的分组比较。

### 2.3.1 比较分布

通常情况下,产生某一随机变量的统计分布被假定为在各个组中都是相同的。但是这并不一定是事实。例如,在某一组中的随机变量背后的分布可能是正态的,而在另一组中分布则可能是对数正态的。在运用广义线性模型时,两个不同组所产生的随机变量的分布服从指数族分布中的不同形式,这在理论上也是可能的。或者,在结构方程模型中,在某一组中数据可能满足多元正态分布假定,而在另一组数据中此假定必须放宽才行。

有时候,两个或多个经验分布之间的差异对于研究者来说是很有意思的,因此需要进一步的考察。在第 4 章中我们将探讨相对分布方法,这是一种非参数方法,它可以捕捉到经验分布中的差异。

### 2.3.2 比较数据结构

如果研究者在各组中找不到差异,则可能这种一致性已存在于基本的数据结构中。例如,在结构方程模型中进行多组分析时,我们在两组的模型参数间找不到差异,可能只是因为各组的相关矩阵非常相似。反过来说,数据结构上的差异可能表明了参数存在差异,虽然它并不能证明参数存在差异。在涉及列联表的分析中,抽到的零单元格的模式可能在两组间如此不同,致使模型差异的基础已经奠定。

### 2.3.3 比较模型结构

另外一个比较层次是模型。可能变量间的关系在这组中用这种类型代表最好,而在那组中用那种类型代表最好。在广义线性模型中,有人可能会愿意考虑如下情况:两个组有不同的联系函数,但联系函数依赖于同一分布。这表明其中一个为典范联系函数(canonical link function),而另一个则不是。在结构方程模型中,数据分析者可能为 A 组设定出一组结构关系,而为 B 组设定出不同的另一组关系。在包含类别潜变量的模型中,两向的边缘分布在不同的组中可以用不同类型的关联来建模(正如在对数线性的传统中一样)。

### 2.3.4 比较模型参数

即使在最常见的比较类型中也可以有一些变化。通常情况下,我们比较各组中的参数或者自变量的效应。如果潜变量存在,我们还可以检验各组的测量模型中的参数是否相同。因此,在一个模型中,测量参数、结构参数、测量方程中的截距、模型中潜变量的均值(或概率)都可以进行分组比较。进行哪种比较或者哪些比较,则由研究者来决定,只要他的目标不是比较模型中存在的所有参数。后面章节中的示例会演示出很多进行比较的可能的方式。

一个与此相关的问题是比较各组中响应变量的高低(level)。虽然传统上并不用参数比来处理这一问题,但是我们可以把这一问题看做对各组之间的截距或者残余均值(residual means)的对比。这样跨组比较响应变量的高低或均值则与对比率的标准化和分解方法密切相关,进而更为一般地与任何度量变量(metric varaibles)的标准化和分解方法密切相关。

第 3 章

# 线性模型中的比较

Comparison in Linear Models

## 3.1 导 论

在本章中,我们主要研究有如下形式的线性模型:

$$y = X\beta + e \qquad \sim N(0, \sigma^2 I) \tag{3.1}$$

其中,y 和 e 是结果和误差的随机向量,向量规模(size)都为 $N \times 1$,X 为包含固定变量(fixed variable)的 $N \times k$ 的矩阵,β 是规模为 $k \times 1$ 的参数向量,I 为单位矩阵。虽然我们也会考察其他相关的比较方法,包括回归分解,但非参数模型则不在这一章讨论之列。

β 的 OLS 估计量由下式给出:

$$\hat{\beta} = (X'X)^{-1}X'y \tag{3.2}$$

或者,$\hat{\beta}$ 也可以通过 ML 估计获得。对于线性模型(3.1)来说,ML 估计值与式(3.2)的估计值相同。

这种线性模型设定了三种一般性统计方法:方差分析(以下简称 ANOVA)、协方差分析(以下简称 ANCOVA)和多元线性回归。在 ANOVA 中,X 只包括离散变量;在 ANCOVA 中,X 包括离散变量和观察到的协变量。在这两种情况中,X 常常被称为设计矩阵(design matrix)。在多元回归中,X 包括预测或解释结果的自变量。这些解释性变量可以是类别型的,也可以是数量型的,或者两者并存。ANOVA 和 ANCOVA 的主要目标是比较各组的均值;这也可以作为线性回归的目标之一。

本章我们只考虑分组的固定效应模型。这样的模型中的解释因子被视为是非随机的,β 是有待估计的非随机系数的向量。随机效应和随机系数模型将在第 10 章中考察。此外,我们将主要考虑 y 中的元素是独立同分布(以下简称 i. i. d. )的情形。此外,我们还将推后讨论由于重复测量或整群测量而导致的 y 彼此相关问题。同样,我们也把比较几率比(comparing odds ratio)的讨论推后到有关广义线性模型的一章中。最后,虽然本章我们主要使用的是一

个简单的单因素示例,但这个例子可以推广至多元和交互因素的模型中。

## 3.2　示　例

表 3.1 中的数据给出了在某一医院出生的 24 位婴儿的“出生体重(g)”、“胎龄(周)”、“性别”。男婴和女婴的平均胎龄很相近(男性为 38.333,女性为 38.750)。人们对这些数据至少可以提出三个问题。人们可能想了解平均出生体重是否有性别差异,同等胎龄情况下平均出生体重是否有性别差异,男婴和女婴的发育速度是否相同。这些问题可以通过一个线性模型来分析数据从而得到解答。

假定发育速度是线性的,则有式(3.1)的特定模型如下:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i \qquad (3.3)$$

其中,$y_i$ 为第 i 个婴儿的出生体重,$x_{1i}$为性别(男性为 1,女性为 0),$x_{2i}$为胎龄,$e_i$为服从 $N(0, \sigma^2)$的独立同分布的随机误差。式(3.3)中的参数估计值就是第 2 章中讲的要比较的检验统计量 $s_g$,检验参数估计值将使我们能够回答上述问题,但首先我们要检验的是 $\beta_2$ 和 $\beta_3$ 的某一个或两个同时被约束为零的模型。

**表 3.1　分性别的出生体重与妊娠周数**

| 男性 | | 女性 | |
|---|---|---|---|
| 出生体重 | 胎龄 | 出生体重 | 胎龄 |
| 2 968 | 40 | 3 317 | 40 |
| 2 795 | 38 | 2 729 | 36 |
| 3 163 | 40 | 2 935 | 40 |
| 2 925 | 35 | 2 754 | 38 |
| 2 625 | 36 | 3 210 | 42 |
| 2 847 | 37 | 2 817 | 39 |
| 3 292 | 41 | 3 126 | 40 |
| 3 473 | 40 | 2 539 | 37 |
| 2 628 | 37 | 2 412 | 36 |
| 3 176 | 38 | 2 991 | 38 |
| 3 421 | 40 | 2 875 | 39 |
| 2 975 | 38 | 3 231 | 40 |

资料来源:Dobson(1990, 表 2.3)。

## 3.3 一些初步考察

第一个假设是关于出生体重相同的原假设：

$$H_0: \mu_1 = \mu_2$$

其中，$s_g$ 由 $\mu_1$ 和 $\mu_2$ 来代表，分别表示男婴和女婴分组的平均出生体重。男婴的样本均值为 3 024，女婴的样本均值为 2 911.333。

在下一章中，我们会用一些非参数方法比较这些数据；现在则主要考虑参数方法。我们做出如下假定，以便从在此考察的 ANOVA 和其他一些参数多重比较模型中进行推论：

- 所有的 $e_i$ 都服从正态分布。
- 所有的 $e_i$ 都是独立分布的。
- 在不同的组中，$e_i$ 的方差 $\sigma_e^2$ 都相同。

许多传统的同方差性检验，如巴利特检验（Bartlett's）、哈特利检验（Hartley's）以及科克伦检验（Cochran's）都基于正态理论，而且对于正态性假定非常敏感；它们对于长尾分布（long-tailed distributions）会得到太过显著的结果。近年来，发展的检验方法并不敏感，更适用于诸多分布类型。例如，莱文（Levene，1960）提出使用独立同分布的正态得分 $z_{ig} = (y_{ig} - \bar{y}_{.g})^2$ 进行 F 检验。其他转换得分也被加以考察：$z_{ig} = |y_{ig} - \hat{y}_{.g}|$，$z_{ig} = \ln|y_{ig} - \bar{y}_{.g}|$，以及 $z_{ig} = |y_{ig} - \bar{y}_{.g}|^{1/2}$。转换得分间存在显著差异被视为 $\sigma_g^2$ 中存在显著差异的证据。对于出生体重数据来说，莱文检验给出的 F 值为 0.029，概值为 0.865，表明方差相等。

## 3.4 线性模型

我们第一个模型是式（3.3）的简化形式：

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i \tag{3.4}$$

在此，我们不关心胎龄对于发育的影响或者发育在不同性别间的差异。我们想回答的唯一问题是男婴和女婴的出生体重是否有差异。如果从分组的角度来看每个变量，则 $\beta_0$ 代表女婴的平均出生体重（估计出的 $\mu_1$），$\beta_1$ 给出了男婴与女婴的差异。因此，$\mu_2$ 可用 $\Sigma_g \beta_g$①（g = 0，1）来估计。$\beta_0$ 和 $\beta_1$ 的普通最小二乘（以下简称 OLS）估计值分别为 2 911.333 和 112.667。但是 $\beta_1$ 在统计上并不显著（t = 0.977，p = 0.339）。

---

① 原文为 $\Sigma_g \mu_g$，疑为笔误。——译者注

## 3.5　比较两个均值

在同方差性假定下(即 $\sigma_1=\sigma_2$),恰当的检验方法是用 t 检验来比较两个均值,在每一本统计学入门教材中都会论及这一问题。大多数统计软件包都会区分单样本比较和双样本比较(或独立样本比较)。单样本程序比较要考察的均值和一些假设的位置取值;而双样本程序比较从两个独立样本得到的位置参数,如均值。这里,以及在第 4 章中处理非参数检验时,我们主要关注双样本比较(或独立样本比较)。那么

$$t_{[n_1+n_2-2]}=\frac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{\hat{\sigma}_{\bar{\mu}_1-\bar{\mu}_2}}=\frac{\bar{y}_1-\bar{y}_2}{\hat{\sigma}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \tag{3.5}$$

在此,均值无差异的原假设被加以检验,而其中:

$$\hat{\sigma}=\sqrt{\frac{(n_1-1)\hat{\sigma}_1^2+(n_2-1)\hat{\sigma}_2^2}{n_1+n_2-2}}$$

而 $n_1+n_2-2$ 是与检验相关的自由度数目。使用表 3.1 中的数据,我们给出分性别对均值和分布特征的图形比较(见图 3.1);这样我们得到的 t 统计量为 0.977,自由度为 22($p=0.339$)。很明显,结果与线性模型中得到的完全相同。尽管从表面看男婴比女婴平均重 113 g,但是体重相等的假设并不能在统计上被拒斥。

一个与此相关的问题是检验两个比例之间的差异。对于这种检验,我们可以很容易地得到式(3.5)的一个变化形式:

$$t_{[n_1+n_2-2]}=\frac{p_1-p_2}{\sigma\sqrt{\dfrac{p_1(1-p_1)}{n_1}+\dfrac{p_2(1-p_2)}{n_2}}} \tag{3.6}$$

其中,$p_g$ 为第 g 组中观察到的事件比例,$p_g(1-p_g)$ 给出其方差。在比较比例或比率时,要比较的组的成分结构常常会影响到比较。因此,在进行比较前,则有必要对某一组(有时对两组)进行标准化。我们将在第 5 章考虑这一问题。

## 3.6　方差分析

单因素 ANOVA 模型同样使用式(3.4),常用于评估处理效应(treatment effects)。在此我们想知道响应变量的均值在 G 组中是否有差异。有时,研究者可以依据某一因素的水平随机分配实验单元给处理组和对照组。这种程序被称为完全随机化的实验设计。在社会科学中,这种设计基本上是不可行的,

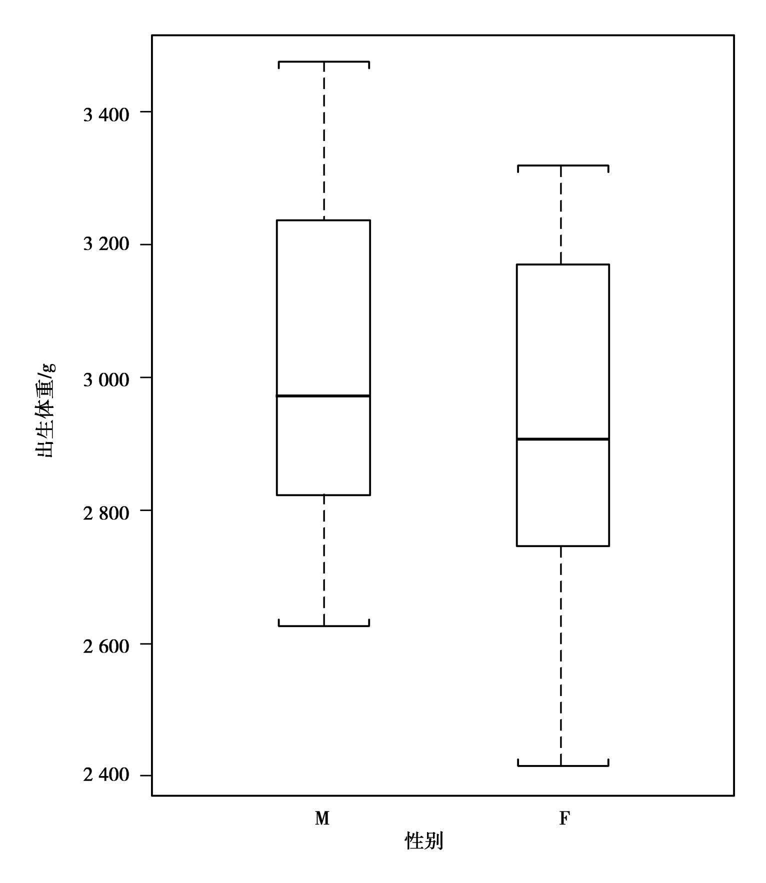



图 3.1　出生体重数据的箱图

而是使用准实验或者非实验设计。不管怎样,如果潜在的混杂因素已经得到控制,那么 ANOVA 就可以被用以评估某一结果在各组间的差异。

ANOVA 中的关键概念是组间平方和($SS_B$)、组内平方和($SS_W$)和总平方和($SS_T$)。它们的定义依次为:

$$SS_T = \sum_{g=1}^{G} \sum_{i=1}^{n_g} (y_{ig} - \bar{y}_{..})^2 \tag{3.7}$$

$$SS_W = \sum_{g=1}^{G} \sum_{i=1}^{n_g} (y_{ig} - \bar{y}_{g.})^2 \tag{3.8}$$

$$SS_B = \sum_{g=1}^{G} n_g (\bar{y}_{i.} - \bar{y}_{..})^2 \tag{3.9}$$

其中,$n_g$ 代表第 g 组中的个案总数。这三者之间的关系由下列恒等式给出:

$$SS_T = SS_B + SS_W \tag{3.10}$$

相关的自由度分别为:$SS_B$ 为 $G-1$,$SS_W$ 为 $N-G$,$SS_T$ 为 $N-1$。将 $SS_B$ 和

$SS_W$ 除以它们各自的自由度数目，便可得到它们的平均平方和（mean sums of squares）①，$MS_B$ 和 $MS_W$。使用 $MS_B$ 和 $MS_W$，接下来我们可以定义三个更为有用的概念。方差成分 $\sigma_e^2$ 和 $\sigma_\alpha^2$ 的无偏估计量分别由下面两式给出：

$$\hat{\sigma}_e^2 = MS_W \quad (3.11)$$ ②

$$\hat{\sigma}_\alpha^2 = (MS_B - MS_W)n_0 \quad (3.12)$$

其中，$n_0 = \left(N^2 - \sum_{g=1}^{G} n_g^2\right) / N(G-1)$。如果各组大小相等，即 $n_1 = n_2 = \cdots = n_g$，我们有 $n_0 = n_g$。从 $\sigma_e^2$ 和 $\sigma_\alpha^2$ 的估计量中，可以得到总方差 $\sigma_y^2$ 的估计如下：

$$\hat{\sigma}_y^2 = \hat{\sigma}_e^2 + \hat{\sigma}_\alpha^2 \quad (3.13)$$

方差分析的 F 检验被定义为 $MS_B/MS_W$，自由度由 $SS_B$ 和 $SS_W$ 定义。对于出生体重的数据，即得到 $MS_B$ 为 76 162.7，$MS_W$ 为 79 714.1，由此给出 $F_{[1,22]}$ 的比值为 0.955，概值为 0.339。

## 3.7　多重比较方法

### 3.7.1　最小显著差异检验

这种程序由费希尔（Fisher，1935）提出，包括两个步骤：

1. 首先，进行方差分析，以确定各组间是否存在足够显著的差异，这要通过在某一选定的显著水平 α 下的整体 F 检验表明。

2. 如果 F 检验显著，就用 t 检验对所有组进行两两比较（pairwise comparison）。

在进行比较时，最小显著差异（以下简称 LSD）统计量被作为标准：

$$|\bar{y}_g - \bar{y}_h| > t_{[N-G,1-\alpha/2]} \sqrt{MS_W\left(\frac{1}{n_g} + \frac{1}{n_h}\right)} \qquad 其中\ g \neq h \quad (3.14)$$

LSD 在不等式右端给出。研究者比较第 g 组和第 h 组的样本均值的观察差异与 LSD 值。如果样本差异大于 LSD，人们就可得出结论，认为差异显著。

表 3.1 中的数据给出的样本差异是 112.667，而数据的 LSD 估计为：

$$t_{[22,1-0.05/2]} \sqrt{MS_W\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 2.074 \sqrt{\frac{79\,714.1}{6}} = 239.057$$

结果显示，差异小于 LSD，并不显著。

### 3.7.2　图基法

图基法（Tukey's Method）是用来比较均值 $\mu_g$ 和 $\mu_h$（$g \neq h$）的所有可能的两

---

① 也常被称为“均方”。——译者注

② 原文为 $= SS_W$，疑为笔误。——译者注

两差异的一种联立检验(a simultaneous test)(Tukey,1953)。目标是为所有两两比较构建 $100(1-\alpha)\%$ 的置信区间。

这种方法的基础是学生化的极差分布(studentized range distribtution)。令 $r$ 为从均值为 $\mu$、方差为 $\sigma^2$ 的正态分布中得到的一些独立观察的极差(即最大值减最小值),令 $\nu$ 为误差的自由度数目(多重比较中为 $N-G$)。则学生化的极差可以定义如下:

$$q_{G,\nu} = \frac{r}{\hat{\sigma}} \tag{3.15}$$

这样,图基法就给出了如下的均值比较的置信区间:

$$\bar{y}_g - \bar{y}_h \pm \frac{1}{2}\sum_{g=1}^{G} |c_g| q_{[G,N-G,1-\alpha]} \sqrt{MS_W \frac{1}{n_g}} \tag{3.16}$$

其中,$c_g$ 是 G 组中的任意比对(arbitrary contrasts),通常有约束条件为 $\sum_{g=1}^{G} c_g = 0$。图基法原本设计为比较两个均值 $\mu_1$ 和 $\mu_2$,故在"±"号后的第一项就成为:

$$\frac{1}{2}\sum_{g=1}^{G} |C_g| = \frac{1}{2}(|1| + |-1|) = 1$$

虽然以上方法要假定各组容量相等,图基(Tukey,1953)和克雷默(Kramer,1956)也发展出一种修正后的检验,其中在式(3.16)中用 $n_g$ 和 $n_h$ 的谐和均值替换了 $n_g$。在许多统计教材中都有 $q$ 的分布。例如,对于表 3.1 中的数据,$q$ 的第 95 百分数值为 $q_{[2,22;0.05]} = 2.94$。这表明,如果我们从正态分布中抽取两个观察,它们的极差不超过自由度为 22 的独立样本标准差估计值的 2.94倍的概率为 0.95。

$$\frac{1}{2}\sum_{g=1}^{G} |c_g| q_{[2,22;1-0.05]} \sqrt{MS_W \frac{1}{n_g}} = (1)2.94 \sqrt{\frac{79\ 714.1}{12}} = 239.621$$

因此,在不同性别间出生重量的差异的 95% 置信区间为 112.667 ± 239.621。

### 3.7.3 谢弗法

谢弗法(Scheffé's method)也为多重比较构建一个 $100(1-\alpha)\%$ 的联立置信区间(Scheffé,1953,1959)。区间由下式给出:

$$\bar{y}_g - \bar{y}_h \pm f \sqrt{(G-1) MS_W \sum_{g=1}^{G} c_g^2 \frac{1}{n_g}} \tag{3.17}$$

其中,$f^2 = F_{[G-1,N-G;1-\alpha]}$,表示自由度为 $G-1$ 和 $N-G$ 的 F 分布的 $100(1-\alpha)$ 百分数点。对于目前的数据,我们使用式(3.17)可得到 95% 的置信区间范围为:

$$\pm \sqrt{f^2 (G-1) MS_W \sum_{g=1}^{G} c_g^2 \frac{1}{n_g}} = \pm \sqrt{4.30(1)79\ 714.1(2)\frac{2}{12}} = \pm 338.019$$

谢弗法更具有普适性,因为所有可能的对比都可用它来检验统计显著性,

而且可为参数的相应线性函数构建置信区间。

作为两种主要的多重比较方法，图基法和谢弗法各有其优缺点，总结如下（Sahai and Ageel，2000：77）：

1. 谢弗法可应用于样本量不等时的多重比较，而原始的图基法只适用于样本量相同时的比较。

2. 在比较简单成对差异（simple pairwise differences）时，图基法最具效力，给出更窄的置信区间，虽然它对于广义比对（general contrasts）也可适用。

3. 与此相比，对于涉及广义比对的比较，谢弗法更具效力，给出更窄的置信区间。

4. 如果 F 检验显著，那么谢弗法将从所有可能的比对（contrasts）中至少检测出一对比对是统计显著的。

5. 谢弗法应用起来更为方便，因为 F 分布表比图基法中使用的学生化极差分布更容易得到。

6. 正态性假定和同方差性假定对于图基法比对于谢弗法更加重要。

### 3.7.4　邦弗朗尼法

邦弗朗尼法（Bonferroni method）适用于在 ANOVA 中事先选定的两两比对或者线性组合的集合，各组样本量既可以相同也可不同。这一集合不像在谢弗法中那样是无限的，但可以超过图基法中的比较集合。这种方法基于邦弗朗尼不定式，它的表述是：如果存在 M 个零假设，每个零假设在 $\alpha/M$ 水平下被拒斥，那么总的误差率最多也就是 $\alpha$。同样，如果在 $100(1-\alpha/M)\%$ 水平下存在 M 个置信区间，那么它们这些置信区间将至少在 $100(1-\alpha)\%$ 水平下同时为真。用数学语言来表达这一原则，假定每个假设 m 在 $\alpha^*$ 水平下受检验，被拒斥的概率为 $R_m$；那么犯第一类错误的总概率为：

$$\alpha = \prod_{m=1}^{M} R_m \leqslant \sum_{m=1}^{M} R_m = M\alpha^* \qquad (3.18)$$

这一程序非常简单，只要 M 不是非常大就可以给出很合理的结果，但对大于 $1-\alpha$ 的真置信区间趋于比较保守。如果只进行少量比较（少于 $G^2$），那么这种方法就更为适用（Fleiss，1986）。在本数据的例子中，由于只有一个零假设，因此总的误差率与单个假设的误差率是相同的。

## 3.8　ANCOVA

ANCOVA 最早是由费舍尔引入的（Fisher，1932），它将 ANOVA 和回归合在一起来评估实验组之间均值差异的统计显著性，同时要根据一个或多个伴随变量（协变量）的线性效应来调整依变量的初始观察值。涉及一个协变量的最简单的 ANCOVA 包括以下三个计算步骤：我们先计算伴随变量的平方

和，计算依变量的平方和，计算两者乘积的平方和。有关这一问题的数学陈述，请参看谢弗的著作（Scheffé，1959）。

ANCOVA 的一个基本假定是各组的回归系数的方差具有同质性，这一假定保证了可以对伴随变量的任何值进行调整。如果没有这一假定，那么调整的效果将因协变量取值的不同而有所不同。如果回归系数并不具有方差同质性，则可用约翰逊-内曼程序（Johnson-Neyman procedure）处理。

例如，研究者想了解一种新英语教材对于学生阅读能力的效果，一组学生使用新英语教材，而另一组学生使用一种标准教材。我们完全有理由怀疑“社会经济地位”将与“阅读能力”有关。“社会经济地位”用学生父亲的职业声望得分来测量。为了对这个协变量进行调整，首先我们要计算这个协变量与“阅读能力”之间的相关，从中可以估计出“阅读能力”的方差中有多少是用“社会经济地位”可以解释的，还有多少残余方差。这些残余方差再被用于 ANOVA 中，来估计“社会经济地位”得到控制之后的真正的 $SS_B$。

在出生体重数据中，ANCOVA 模型给出的 $MS_B$ 为 157 304，$MS_W$ 为 31 370，由此得到的 F 统计量为 5.014（$p=0.036$）。因此，尽管两性之间的胎龄似乎是相等的，但一旦控制了胎龄的效应后，两性之间出生体重的差异就显著了。

## 3.9　多元线性回归

我们现在将式（3.4）进行扩展，使其包括变量 $x_2$（胎龄）：

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \tag{3.19}$$

如果 $x_2$ 保持不变，$y_i$ 的差异在 $\alpha=0.05$ 水平下变得统计显著了，$\hat{\beta}_1=163.039$，其标准误为 72.808（$p=0.036$）。结果当然与前面 ANCOVA 的结果完全相同（注意概率水平完全相同）。因此，控制了“胎龄”之后，估计的出生体重的性别差异大于观察到的 112.667。“胎龄”的影响是高度显著的（$\hat{\beta}_2=120.894$，标准误为 20.463），表明存在每周大约增加 121 g 的线性增长率。结果，模型的 $R^2$ 也从式（3.4）的 0.042 增加到了 0.640。

至此我们已经发现，数据不支持在第 3.2 节开始时提出的第一个假设，但是支持第二个假设。现在转向第三个假设：“胎龄”的影响是否有性别差异？这一问题可以通过检验形式式（3.3）的模型得到解答。但是数据并不支持“胎龄”的影响有性别差异，对于 $\beta_3$ 的 t 检验（$t=-0.441$，$p=0.664$）和对 $R^2$ 的增加量的 F 检验（$F_{1,20}=0.168$）都显示了这一点。

另一种检验包含了交互项（它代表了分组变量的效应差异）的模型的方法是邹氏检验（Chow，1960）。这种使用非常普遍的方法是用来检验两个（或多个）回归之间的差异，它基于以下假定：

- $e_{1i} \sim N(0, \sigma^2)$

- $e_{2i} \sim N(0, \sigma^2)$
- $cov(e_{1i}, e_{2i}) = 0$

也就是说,回归 1 和回归 2 的误差项服从均值为零且有着相同方差 $\sigma^2$ 的正态分布,并且它们彼此独立。在满足如上假定的前提下,邹氏检验按如下步骤进行:

1. 运行样本量为 N 的合并样本的回归方程(在本例中,$n_1 + n_2 = N = 24$),得到 $SS_W$,称其为 $S_1$。自由度为 $n_1 + n_2 - k$,其中,k 为要估计的参数数目(包括截距项)。

2. 分别运行样本量为 $n_1$ 和 $n_2$ 的分组样本的回归方程,得到它们的 $SS_W$,或者 $S_2 + S_3$,自由度为 $n_1 + n_2 - 2k$。

3. 进行如下定义的 $F_{[k, n_1 + n_2 - 2k]}$ 检验:

$$F = \frac{[S_1 - (S_2 + S_3)]/k}{(S_2 + S_3)/(n_1 + n_2 - 2k)} \tag{3.20}$$

如果得到的 F 统计量超过了 F 的临界值,我们就拒斥两个回归相同的零假设。将邹氏检验应用于出生体重数据中,我们得到:

$$F = \frac{[816\ 074.4 - (403\ 698.7 + 248\ 725.9)]/2}{(403\ 698.7 + 248\ 725.9)/(12 + 12 - 4)} = 2.508$$

$F_{2,20}$在 $\alpha = 0.05$ 时的临界值为 3.49。即使在 10% 水平下上述结果也并不显著(在 10% 水平下 $F_{2,20}$的临界值为 2.97)。因此,我们既不能断定在控制了"胎龄"后两性的"出生体重"存在差异,也不能断定"胎龄"对于体重增长的效应在两性之间有所不同。图 3.2 显示了两性各自的拟合回归线和观察到的数据点。从中可以看出,两性的回归线斜率间的差异是很小的。

或者,我们也可用"性别"和"胎龄"的交互作用来检验组间差异,如式(3.3)所示。这一思想也适用于组数超过两组、自变量不止一个的更为一般的情况中。

与多元线性回归有关的一种比较方法被称为多元分类分析(以下简称 MCA)。20 世纪 60 ~ 70 年代是这种方法的黄金时代,但之后便不再流行。在 MCA 中,解释变量都是类别型的,然后对其进行"效应编码"(effect-coded)来显现它们与总均值的偏离程度。在简洁明了地展现比较结果方面,这种方法仍然非常有用。例如,哥尔舒尼(Gershuny,2000)最近分析了 20 个国家的 35 项调查数据,时间跨度从 20 世纪 60 ~ 90 年代,介绍了世界范围内时间分配的变动情况(如花在有酬工作、家务劳动、休闲上的时间等)。

在使用线性回归分析进行组间比较时出现的一个实际问题是各组间在样本量上的差异。前面讨论的邹氏检验考虑到在进行比较时各组自由度的差异。或者,一些研究者对于样本量喜欢采用一种更为实际的方法。例如,在哥尔舒尼(Gershuny,2000)对各个国家和各个时期的时间分配分析中,他对每个国家的个案数进行加权,加权后产生的总数都是 2 000,男女分别各占 1 000

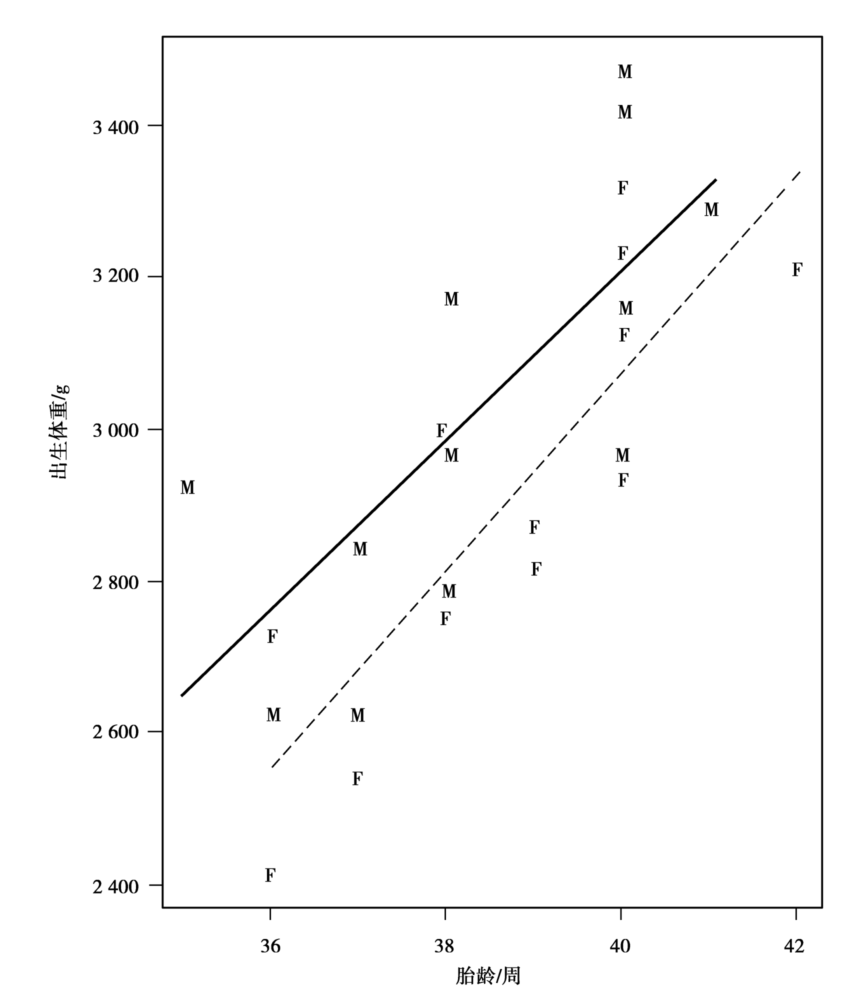



图 3.2　出生体重与胎龄的带有回归线的示意图

左右。这类加权还可同其他加权方案结合使用。在哥尔舒尼的时间利用研究中,他对于每个国家每个时间点的样本都进行了加权,以使这些样本与 1961 年的英国(即英国调查的第一个时间点)的社会经济成分有可比性,这样便产生了“英国基准年加权”分析。读者可能会发现,这种依据成分组成来加权的方法非常接近于我们将在第 5 章中介绍的标准化方法。

# 3.10　回归分解

## 3.10.1　基本原理

回归方法的最通常的用处之一就是解释为什么某种社会的、经济的、心理的或医学的结果在不同组中有差异。例如，在研究性别差异时，通常我们想确定两性在社会网络、在工资率、在心情压抑、在对某种治疗的反应上差异的成分(components)。返回到出生体重的示例上，我们会考虑两种彼此竞争的解释或假设：

- 男孩在胚胎阶段需要更长的时间来成熟，因此出生时更重。
- 女孩从遗传上讲身高就低一些，因此体重更轻。

这两个假设并不互斥；它们可能同时为真。为了评估出生体重的差异中有多少可以用这些假设来解释，我们转向回归的分解。

## 3.10.2　代数表达式

我们把式(3.19)的多元回归方程重新整理，用下标而不用变量的形式来表示组：

$$y_{gi} = \beta_{0g} + \beta_{1g} x_{1gi} + e_{gi} \tag{3.21}$$

其中，g 表示两个性别组，1 和 2。我们分三个代数步骤来分解组均值的差异。首先，我们使用 $y_g$ 和 $x_{1g}$ 的分组均值拟合过原点的回归线。这样，回归线就穿过了两个性别组的样本均值点(1 = 男；2 = 女)：

$$\bar{y}_1 = \hat{\beta}_{01} + \hat{\beta}_{11} \bar{x}_{11} \tag{3.22}$$

$$\bar{y}_2 = \hat{\beta}_{02} + \hat{\beta}_{12} \bar{x}_{12} \tag{3.23}$$

接下来，我们用第 1 组的均值减第 2 组的均值，得到下式：

$$\bar{y}_1 - \bar{y}_2 = (\hat{\beta}_{01} - \hat{\beta}_{02}) + (\hat{\beta}_{11} \bar{x}_{11} - \hat{\beta}_{12} \bar{x}_{12})$$

最后，我们再加上和减去 $\hat{\beta}_{11} \bar{x}_{12}$，重新整理各项后得到下式：

$$\bar{y}_1 - \bar{y}_2 = (\hat{\beta}_{01} - \hat{\beta}_{02}) + \bar{x}_{12}(\hat{\beta}_{11} - \hat{\beta}_{12}) + \hat{\beta}_{11}(\bar{x}_{11} - \bar{x}_{12}) \tag{3.24}$$

我们现在将两组均值的差异表达为以下三项之和：

$$\hat{\beta}_{01} - \hat{\beta}_{02}, \quad \bar{x}_{12}(\hat{\beta}_{11} - \hat{\beta}_{12}) \text{ 和 } \hat{\beta}_{11}(\bar{x}_{11} - \bar{x}_{12})$$

## 3.10.3　解释

第一项 $\hat{\beta}_{01} - \hat{\beta}_{02}$ 给出的是常数项或截距项的差异。第二项 $\bar{x}_{12}(\hat{\beta}_{01} - \hat{\beta}_{02})$ 给出的是参数估计的差异，再用第 2 组中 $x_1$ 的均值加权。第三项 $\hat{\beta}_{11}(\bar{x}_{11} - \bar{x}_{12})$

给出的是解释变量均值的差异，再用第 1 组的参数估计加权。表 3.2 中报告了这些组成部分的估计值。

表 3.2　对于出生体重数据的回归分解的各组成部分

| | |
|---|---|
| $\bar{x}_{11}$ | 38.33 |
| $\bar{x}_{12}$ | 38.75 |
| $\hat{\beta}_{01}$ | −1 268.67 |
| $\hat{\beta}_{02}$ | −2 141.67 |
| $\hat{\beta}_{11}$ | 111.98 |
| $\hat{\beta}_{12}$ | 130.40 |

这些组成部分能够告诉我们什么？出生体重的差异被分解为三个部分。$\hat{\beta}_{11}(\bar{x}_{11}-\bar{x}_{12})$测量的是由于胎龄上的性别差异而导致的出生体重上的性别差异量，它是从胎龄的“标准”效应方面来表示的（在本例中是男孩的胎龄的效应）。这一项评估了前面所述的“成熟假设”。$\bar{x}_{12}(\hat{\beta}_{01}-\hat{\beta}_{02})$表示了由于胎龄的效应的差异而导致的出生体重方面的性别差异，它是从女孩的平均胎龄方面来进行估值的。这一项评估了前面所述的“遗传假设”。最后，$\hat{\beta}_{01}-\hat{\beta}_{02}$给出了出生体重的未能解释的残余差异。如果回归模型并不存在被忽略的变量导致的设定错误，那么，$\bar{x}_{12}(\hat{\beta}_{01}-\hat{\beta}_{02})+(\hat{\beta}_{01}-\hat{\beta}_{02})$就概括了遗传方面的差异。有人可能会注意到，分解并不是唯一的，因为如果用$\hat{\beta}_{12}\bar{x}_{11}$来代替$\hat{\beta}_{11}\bar{x}_{12}$，那么，第 2 组就会被视为比较的标准组或基准组（它的胎龄增加率将会被使用）。

出生体重的平均差异是3 024.00 −2 911.33 =112.67，它现在可以被分解为三部分。对于“发育假设”来说，$\hat{\beta}_{11}(\bar{x}_{11}-\bar{x}_{12})=-47.03$，这显示了由于胎龄上的性别差异而导致的平均出生体重上的性别差异量，它是从男孩的胎龄的“标准”效应方面来进行估值的。对于“遗传假设”来说，$\bar{x}_{12}(\hat{\beta}_{11}-\hat{\beta}_{12})=-713.78$，这给出了由于胎龄的效应差异而导致的平均出生体重方面的性别差异，它是从女孩的平均胎龄方面来进行评估的。残余效应为$\hat{\beta}_{01}-\hat{\beta}_{02}=873.00$，这给出了在平均出生体重上未能解释的差异。这三个组成部分加总为 112.19，而平均体重差异为 112.67，这是由于四舍五入的误差造成的。总而言之，遗传对于两性的体重差异的贡献几乎是发育时间的贡献的 15 倍，但是仍然存在许多未能解释的效应。

### 3.10.4　扩展到多元回归

到目前为止，我们还只是对包含一个解释变量的回归模型的分解进行了一番考察。我们可以将它推广到有着多个自变量的回归模型中，则式(3.24)

变为：

$$\bar{y}_1 - \bar{y}_2 = (\hat{\beta}_{01} - \hat{\beta}_{02}) + \sum_{k=1}^{K} \bar{x}_{k2}(\hat{\beta}_{k1} - \hat{\beta}_{k2}) + \sum_{k=1}^{K} \bar{\beta}_{k1}(\bar{x}_{k1} - \bar{x}_{k2}) \quad (3.25)$$

其中，下标 k 表示解释变量 x 的序号，K 表示 x 的总数。对于涉及多个组的回归分解来说，一种比较自然的方法是进行成对比较，这样式(3.24)仍然适用。

## 3.11　应该使用哪一种线性方法

在本章中，我们讨论的许多比较方法虽然都是基于线性模型的，但是它们的统计基础却彼此不同。在进行比较研究时，我们究竟应该使用哪一种方法呢？

这时我们应该考虑的第一个问题是研究设计是否是实验的。如果是随机化已经避免了任何潜在混杂因素的实验设计，那么，用一个简单 t 检验就足以进行均值比较了。如果是涉及实验设计的多组比较，那么，可以采用 3.5 节中介绍的多重比较方法中的某一种。在同一节中，还对各种多重比较方法的优劣进行了比较(虽然远不够完整)。

很多情况下真正的实验难以做到，研究采用的是准实验设计或者非实验设计。这意味着很可能存在一些成分因素(compositional factors)或非成分因素混杂在比较中。如果所有的混杂变量都是类别变量，就可以使用 ANOVA。如果存在协变量，就应选择使用 ANCOVA。在此，我们只考察要比较的结果变量只有一个的情形。在存在多个结果的情形时，我们可以采用 MANOVA 和 MANCOVA。

多元回归方法与 ANOVA 和 ANCOVA 密切相关，因此，可以用于希望进行跨组比较的准实验设计和非实验设计中。是选择回归还是选择 ANOVA 和 ANCOVA，在很大程度上是任意的，它与学科传统有关。例如，在心理学和教育研究中，ANOVA 和 ANCOVA 更为流行，而在许多社会学和政治学研究中，使用回归方法更多。虽然多元回归已经被扩展为诸如半相依回归模型(seemingly unrelated regressions)这样的对多个响应变量建模的方法，但是对于多个组之间的比较而言，只要真正掌握了经典的多元回归，我们便足以将它推广运用到同时考虑多个彼此相关的结果的模型中去①。

---

① 半相依回归模型是由泽尔纳(Zellner，1962)提出的统计模型，其中包含有多个方程，各个方程的自变量和依变量都彼此不同，但误差项却彼此相关，故称为半相依回归模型，目前较多地应用于计量经济学研究中。——译者注

# 第4章 非参数比较

Nonparametric Comparison

在本章中,我们考虑几种主要的非参数统计比较方法。这些方法包括柯尔莫哥洛夫-斯米诺夫检验(Kolmogorov-Smirnov test)和曼-惠特尼检验(Mann-Whitney test)的非参数检验法、排列法(permutation methods)、自助法(boot-strapping methods)、相对分布法。虽然并未涵盖每一种非参数方法,但是这些方法代表了那些最常用的传统非参数方法和那些最新的非参数方法,如相对分布法。

虽然所有这些方法在本质上都是非参数的,但是其中一些方法可以用于参数检验中,以进行统计比较。例如,排列法和自助法都可以用于进行检验,如 t检验或 F 检验。尽管可以用于参数检验,但我们还是把它们列入本章中,因为构成它们的理论基础中并无参数方法中常见的有关分布的假定。

## 4.1 非参数检验

有时候参数检验并不合适,因为它们背后的分布假设有问题。例如,双样本 t检验假定样本间的差异服从正态分布,而非参数检验则没有这样的假定。因此,如果正态分布假定不成立,下面的非参数检验就用得上了。此外,非参数检验并不一定要求数量型数据。例如,曼-惠特尼检验可以用于定序的或有序列的非数量型数据。自然地,本节中用于检验的结果变量是多取值的(multivalued)。对于二分结果,我们可以使用麦克尼马尔检验(McNemar test)。

### 4.1.1 柯尔莫哥洛夫-斯米诺夫双样本检验

这是一种对两个分布间的差异的检验,它基于两个样本的相对累积频数分布之差的绝对值(unsigned differences)。其临界值可以在柯尔莫哥洛夫-斯米诺夫统计表中查找,以便确定两个累积频数分布间的最大差异是否显著。检验统计量如下:

$$D = n_g n_h \max \left| \frac{F_g}{n_g} - \frac{F_h}{n_h} \right| \qquad 其中\ g \neq h \tag{4.1}$$

其中,$F_g$ 是第 g 组的累积频数分布,$\frac{F_g}{n_g}$给出了第 g 组的相对累积分布。对于表 3.1 的数据,临界值为 84($\alpha = 0.05$)。$D = (12)(12)0.167 = 24$,这比临界值要小得多,表明两个分布间不存在显著差异。在 Stata 程序中,D 被定义为 $\max |F_g/n_g - F_h/n_h|$,并且给出了相关的概率值(在本例中为 0.996)。一些统计软件(如 SPSS),也可计算柯尔莫哥洛夫-斯米诺夫统计量的 Z 值。在本例中,柯尔莫哥洛夫-斯米诺夫统计量的 Z 值为 0.408,这个值相当小,表明两组之间不存在显著差异。

### 4.1.2　曼-惠特尼 U 检验

曼-惠特尼检验等效于威尔科克森秩和检验(Wilcoxon rank sum test)。对于配对数据或者单样本数据,威尔科克森符号秩检验(Wilcoxon signed rank test)是合适的。对于我们要关注的双样本数据,则可以进行曼-惠特尼检验(或者威尔科克森秩和检验)。这种检验的基础是秩,即测量位置之间的差异。检验包括以下三个步骤:

1. 把两个样本按升序排列。

2. 对于其中一个样本的每个观察(出于方便起见,一般用较小的那个样本,如 g 组),计算在另一个样本中取值比它小的观察的数目。令此数目为 $C_{ig}$(其中,i 为观察在 g 组内排列的序号)。相同取值(ties)以 1/2 计数。

3. 曼-惠特尼统计量 $U_g$ 为 $\Sigma_i C_{ig}$和 $n_g n_h - \Sigma_i C_{ig}$中的较大者。

出生体重数据中得出的 U 值为 58,而临界值为 102($\alpha = 0.05$),这表明两性间出生体重不存在差异。统计表中只给出了 $n_g \leq 20$ 时 U 的临界值。如果 $n_g > 20$,则可按下式计算 t值:

$$t = \frac{\left(U - \frac{n_g n_h}{2}\right) \pm 0.5}{\sqrt{\frac{n_g n_h (n_g + n_h + 1)}{12}}} \tag{4.2}$$

其中,0.5 为连续性修正。如果括号内的项得到的值大于 0,就加 −0.5;否则就加 0.5。这一数量近似服从正态分布,使用∞ 作为其自由度。许多软件都可以产生曼-惠特尼 U 检验;如果样本量较小,我们可以使用下列网址快速得到结果:http://vassun.vassar.edu/ ~lowry/utest.html。

或者,检验统计量也可用下式计算:

$$U_g = n_g n_h + \frac{n_g(n_g + 1)}{2} - T_g \tag{4.3}$$

其中,$T_g$ 为从两组的合并排序中得到的 g 组的观察秩和。对于本例,用手工算出的 U 值为 58,与 SPSS 得出的结果完全相同。

# 4.2 再抽样法

## 4.2.1 排列法

排列检验可以应用于连续数据、定类数据或者类别型数据,数据可以满足也可以不满足正态假定。这种方法非常灵活,因为对于每一种参数检验或非参数检验,都存在一种相对应的分布自由的(distribution-free)排列检验。

检验 $\phi$ 是一种在[0,1]中取值的决策规则。如果 $\phi(x) < p$($p$ 为事先确定的值),我们就拒斥零假设;否则,我们保留零假设。在 $\alpha$ 水平下的排列检验由以下部分组成:一个包括 N 个观察的向量 z,一个统计量 s(z),一个接受标准 A;它们满足以下条件:对于所有的 z 来说,当且仅当满足下式时才有 $\phi(z)=1$:

$$W(z) = \sum_{\pi \in \prod} A[s(z), s(\pi, z)] \leq \alpha N! \tag{4.4}$$

其中,$\prod$ 是 $n_g + n_h$ 个观察的所有的可能重新排列的集合(Good,2000:203)。

作为随机化检验的一个特例,排列检验的逻辑是容易理解的。排列检验可有不同的形式;例如,它可以与蒙特卡洛模拟合并起来使用。它有如下步骤:

1. 选择一个检验统计量 s(z),然后建立它的拒斥规则。
2. 用样本量分别为 $n_g$ 的各个原始组计算样本的 s(z)。
3. 把所有组合并为一个大数据集,样本量为 N。
4. 进行排列:
   (a)对整个样本进行随机排列。
   (b)把得到的样本分为样本量为 $n_g$ 的若干组。
   (c)为新得到的样本计算 s(z)。
   (d)比较新的 s(z)和基于原来样本中得到的 s(z)。
   (e)重复(a)到(d),直到能够做出决策为止。
5. 如果按照 s(z)的排列分布的拒斥规则,原始样本中得到的 s(z)是一个极端值,那么,我们则拒斥零假设;否则,则接受零假设。

如果我们对出生体重数据进行排列检验,可以视婴儿的性别为未知,然后比较我们重新排列后的检验结果与性别已知时原始的检验结果。可以做出一个决策标准,原始检验结果超过某一极端值时,则可以拒斥零假设。或者,我们可以基于从所有排列样本中得到的 s(z)的分布来构建置信区间。用与此相同的逻辑,我们可以基于排列分布来执行 t 检验和其他检验。

为了加深对上述程序的理解,让我们来看一个非常简单的例子,例子中包括了一个结果变量,所有个案分别属于两组。个案的值是用随机正态分布生

成的,四个个案属于第 1 组,两个个案属于第 2 组。第 1 组的均值为 4,方差为 1;第 2 组的均值为 2,方差为 1。在 S-plus 中用矩阵形式表达的数据如下:

```
group outcome
[1,]  1 5.140497
[2,]  1 3.416688
[3,]  1 3.581760
[4,]  1 4.505986
[5,]  2 1.235289
[6,]  2 2.364622
```

使用第 3 章中讨论过的进行均值比较的众多方法中的某一种,如线性回归,我们用 S-plus 计算得到如下结果:

```
> summary(lm(outcome ~ group))
Call: lm(formula = outcome ~ group) Residuals:

    1         2        3        4        5        6
0.9793   -0.7445 -0.5795   0.3448 -0.5647   0.5647

Coefficients:
                 Value Std.  Error t  value   Pr(>|t|)
  (Intercept)       6.5225   0.9885   6.5984   0.0027
       group      -2.3613   0.6990  -3.3782   0.0278
Residual standard error: 0.8071 on 4 degrees of freedom
Multiple R-squared: 0.7405
F-statistic: 11.41 on 1 and 4 degrees of freedom, the
p-value is 0.02783
Correlation of Coefficients:
  (Intercept)
group   -0.9428
```

对于分组变量的参数估计值为 -2.361,这就是两组均值之间的差值(用第 1 组的均值减去第 2 组的均值,可以得到这个值为 -2.361 277)。这个估计值的标准误为 0.699,表明在 3% 的水平下可以拒斥认为组间不存在差异的零假设。

在此,我们的检验统计量 s(z) 为两组均值之差值,但是由于两组的数据已经合并到一个数据集中了,我们可以用排列过程来对其进行处理。在现实中,数据量一般都很大,我们可以按照前面叙述的程序来对排列数据进行随机抽样。但在本例中数据量很小,可使用包括了所有数据排列的精确排列分布

来进行随机化检验。在此，我们生成了一个变量“index”，表示在全部 15 种可能排列中个案所属的组别：

```
> index
      [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]    1    1    1    1    2    2
 [2,]    1    1    1    2    1    2
 [3,]    1    1    2    1    1    2
 [4,]    1    2    1    1    1    2
 [5,]    2    1    1    1    1    2
 [6,]    1    1    1    2    2    1
 [7,]    1    1    2    1    2    1
 [8,]    1    2    1    1    2    1
 [9,]    2    1    1    1    2    1
[10,]    1    1    2    2    1    1
[11,]    1    2    1    2    1    1
[12,]    2    1    1    2    1    1
[13,]    1    2    2    1    1    1
[14,]    2    1    2    1    1    1
[15,]    2    2    1    1    1    1
```

此类排列的逻辑是：个案属于哪个组完全是随机观察到的。因此，原始观察只是这 15 种可能排列中的一种。

为了计算出 s(z) 的排列统计量，首先我们生成了一个长度为 15 的向量，以容纳使用排列分布计算出的 s(z)。然后，为了便于观察，我们将这些统计量排序，再与原始参数检验统计量进行比较：

```
>permstat <- numeric(15)
>for (i in 1:15) permstat[i] <-
+ mean(outcome[index[i,] ==2]) - mean(outcome[index[i,] == 1])
>sort(permstat)
 [1] -2.36127708 -1.57222777 -1.44842381 -0.75525455 -0.72522757
     -0.60142361
 [7] -0.27937135  0.09174565  0.18762570  0.56762884  0.88079497
      1.00459892
[13]  1.35667816  1.48048211  2.17365138
```

原始的检验统计值为 -2.361 277。比较每个 s(z) 与原始统计量，我们只有第一个 s(z) 的绝对值大于或等于原始检验统计值。这表明，在整个排列分布中，只有原始组合的组间均值差值才有那么大，可以算出 1/15 = 0.066 7。如果我们选择传统的 5% 水平，那么，我们并不能够根据排列检验的结果拒斥零

假设。

Stata 软件中有一个用户自己编写的进行排列检验的模块。该软件可以从下列网址获取:http://www.ats.ucla.edu/stat/stata/code/permute.htm. 如果样本量较大,需要进行抽样,那么这一模块是很有帮助的。

## 4.2.2　自助法

自助法(bootstrap)是近年来引入的另一种再抽样方法,它的计算量也非常大。现在已有许多这种方法的应用,也有一些教科书,其中一本很好的导论性的书是由埃弗龙和蒂布希拉尼撰写的(Efron and Tibshirani,1993)。使用自助法的典型检验步骤如下:

1. 选择一个检验统计量 s(z)。
2. 计算原始样本的 s(z)。
3. 对原始样本中每个原始组都进行有放回的再抽样,并重复不断地再抽样,直到达到预定的次数为止,由此得到 s(z) 的自助分布。
4. 比较 s(z) 与自助分布的上 $\alpha$ 百分位数,以决定接受还是拒斥零假设。

与排列检验一样,自助法也适用于任何参数统计量。

理解自助法的关键在于,它并不能帮助我们估计均值、均值差或者其他估计量。它的用处在于帮助我们计算这些估计量的标准误和置信区间,而无须渐进理论,也无须对估计量做出武断的假设。置信区间的四种常见类型有正态的、百分位的、修正偏误的(the bias corrected)和加速的。

基于从自助样本中得到的自助估计量,我们可以获得新的标准误,这样就可以形成相应的正态置信区间:

$$\text{Prob}\left( s(z) \in [s(\hat{z}) - t_{1/2\alpha} \times se(s(\hat{z})), \{s(\hat{z})\} + t_{1/2\alpha} \times se(s(\hat{z}))] \right) = 1 - \alpha$$

其中,$\alpha$ 是犯 I 类错误的概率。很清楚,在进行自助抽样和自助估计后,就应再一次使用正态理论。为了避免这种参数估计中的陷阱,我们可使用百分位置信区间(percentile confidence interval),它依赖于自助估计的经验分布:

$$\text{Prob}\left( s(z) \in [s(\hat{z})^*_{B\times 1/2\alpha}, s(\hat{z})^*_{B\times(1-1/2\alpha)}] \right) = 1 - \alpha$$

其中,$s(\hat{z})^*_{B\times 1/2\alpha}$ 及其上限对应值都是从自助估计的经验分布得出的。例如,我们抽取了 1 000 个自助样本,$s(\hat{z})^*_{B\times 1/2\alpha}$ 由有着第 25 个最低估计值的样本给出,而 $s(\hat{z})^*_{B\times(1-1/2\alpha)}$ 由有着第 25 个最高估计值的样本给出。

经典的自助法中将偏误的估计定义为自助估计中的均值和原始样本的估计值之差值。更为精细的自助法置信区间要考虑到偏误的估计。有关修正偏误的置信区间和加速的置信区间的进一步讨论,读者可参阅埃弗龙和蒂布希拉尼的著作(Efron and Tibshirani,1993)。

下面我们将使用自助法来分析前一节中的数据。为了简化估计,我们使用了线性回归方法,其中组间差异是由虚拟变量的参数代表。大多数主要的

统计软件中至少有一种程序可以使用自助法(如 SPSS 的非线性回归程序,或者 SAS 中的 Proc Multitest 程序),而 Stata 软件中包含了几种非常灵活的自助法命令:bstrap,bs,bstat,bsample,可以将自助法与任何现存的估计方法配合使用。我们使用 bs 来获得自助回归估计:

```
bs "reg outcome group1"  "_b[group1]", reps(20)
```

其中,最后一个括号中的值给出了要求进行自助抽样的次数。表 4.1 给出了自助估计值和置信区间,以及用参数方法得到的回归估计值。看起来,在进行 200 次重复抽样之后状况就再没有任何实质性的改善了,偏误和标准误都不再变小。从有放回进行 20 次抽样中得到的自助估计值的置信区间包括了 0,在 0.05 水平下的 t检验不显著。其他所有置信区间都不包括 0,而我们记得,模拟数据的真正差异为 2。在本例中,自助估计全都给出了更大的标准误,这进一步证实了从排列检验中得到的结果。

**表 4.1 对模拟数据的参数估计与自助估计**

| 样本 | $\hat{\beta}$ | 偏误 | $se(\hat{\beta})$ | 下限 | 上限 |
|---|---|---|---|---|---|
| 原始 | -2.361 | — | 0.699 | -4.302 | -0.421 |
| 20 | -2.361 | 0.568 | 1.151 | -4.770 | 0.048 |
| 50 | -2.361 | 0.286 | 0.806 | -3.981 | -0.742 |
| 100 | -2.361 | 0.310 | 0.984 | -4.315 | -0.408 |
| 200 | -2.361 | 0.128 | 0.844 | -4.026 | -0.696 |
| 500 | -2.361 | 0.200 | 0.832 | -3.996 | -0.727 |
| 1 000 | -2.361 | 0.232 | 0.887 | -4.102 | -0.620 |
| 2 000 | -2.361 | 0.214 | 0.861 | -4.049 | -0.673 |
| 5 000 | -2.361 | 0.205 | 0.866 | -4.059 | -0.663 |

注:所有的置信区间都是 95%。出于简化考虑,表中只列出了正态置信区间。

## 4.3 相对分布法

近年来,对于分布比较的兴趣促进了非参数比较方法的另一种应用。通常研究者感到有必要就社会经济成就、教育机会获得以及与健康有关的行为或态度等问题进行分组比较。分组比较的焦点不限于诸如均值或中位数这样的位置参数。研究者真正感兴趣的是比较分布的形状和整体差异。在本节中,我们考察汉德科克和莫里斯(Handcock and Morris,1998,1999)介绍的相对分布方法。如果读者想了解这一方法的历史源头,可去阅读有关的参考文献。

这一方法关注的是基于相对分布的整体分布比较。多年来,相对分布的概率密度函数(以下简称 PDF)和累积分布函数(以下简称 CDF)已经被独立

地研究了许多年。在这些研究中,CDF 记录一个分布中的得分(如测试结果或收入)对应另一个分布中的百分数;而 PDF 是两个分布的再量纲密度(rescaled density)。沿用汉德科克和莫里斯(Handcock and Morris,1998,1999)的符号,我们用 $F_0(y)$ 表示参照组中结果变量的 CDF。如果我们比较同一组在两个时点的结果,参照组就是起始时点的 CDF。如果我们比较的是两个不同组的结果,研究者可以自行决定把哪个组作为参照组。此外,我们用 $F(y)$ 表示比较组的 CDF。这样,比较组可以是在后一时点上测量的同一组,也可以是同一时点上测量的另一组。

令 $Y_0$ 和 $Y$ 分别是从 $F_0$ 和 $F$ 中抽取的随机变量。$F_0$ 和 $F$ 的 CDF 有着共同的支撑(support)[①],且绝对连续。$Y$ 到 $Y_0$ 的等级变换(grade transformation)定义为以下随机变量:

$$R = F_0(Y) \tag{4.5}$$

其中,$R$ 是$[0,1]$区间上的实数,通过函数 $F_0$ 对 $Y$ 进行变换而得来。随机变量 $R$ 的实现值为 $r$,也被称为相对数据。$R$ 的 CDF 表达式如下:

$$G(r) = F(F_0^{-1}(r)) \qquad 0 \leqslant r \leqslant 1$$

其中,$r$ 表示取值的百分位值,$F_0^{-1}(r)$ 代表 $F_0$ 的分位数函数(quantile function)。$R$ 的 PDF 表达式为:

$$g(r) = \frac{f(F_0^{-1}(r))}{f_0(F_0^{-1}(r))} \qquad 0 \leqslant r \leqslant 1 \tag{4.6}$$

如果两个分布完全相同,则相对分布的 CDF 为一条 45°的直线。相应地,此时相对分布的 PDF 为均匀分布。

相对分布法提供了明确的解释。我们可以把相对数据解释为原本的比较组数值在参照组中的百分数序位,把相对 PDF 即 $g(r)$ 解释为位于结果的给定水平 $F_0^{-1}(r)$[②]上的比较组个案比例与参照组个案比例的密度比。下面有关收入的例子可以说明进行分布比较时解释的一些特点。

在本例中,我们将比较两个欧洲国家——法国和匈牙利。法国的数据来自于 1990 年法国家户调查,是在洛林大区非集体户中收集的;匈牙利的数据基于 1992 年匈牙利家户调查,总体是全国的非集体户。去掉非常少量的缺失值和收入零值后,进行比较的法国样本量为 2 057,匈牙利样本量为 2 030。

首先,我们来用比较收入不平等时最常用的方法——洛伦兹曲线和基尼系数来分析数据。洛伦兹曲线画出的是一个 CDF 曲线,它要回答的问题是:"有着最低收入的那些占人口比例为 $p$ 的人们累积起来共拥有 $Y$ 中的多少呢?"换而言之,洛伦兹曲线比较的是累积收入分布与累积人口分布。如果是绝对平等的,洛伦兹曲线就是一条从(0,0)到(1,1)的 45°的直线。只要是偏

---

① 一个概率分布的支撑是指其补集的概率为零的最小闭集。——译者注

② 原文为 $Y(F_0^{-1}(r))$,疑为笔误。——译者注

离了绝对平等状况,曲线就会处于对角线之下。基尼系数的取值为 0 ~ 1,0 代表完全平等,1 代表绝对不平等。它的取值等于洛伦兹曲线与对角线之间面积的 2 倍。如图 4.1 所示为法国和匈牙利的洛伦兹曲线。法国的洛伦兹曲线稍微更靠近 45°对角线,表明它的收入分配更为平等。同样,法国数据的基尼系数为 0.322,匈牙利数据的基尼系数为 0.369。有许多方法来描述洛伦兹曲线和计算基尼系数,但是最简单的方法是先对数据按 Y 变量排序,然后生成一个 Y 的累积变量和秩变量。之后再分别除以 Y 的最大累积值和个案总数目,由此便得到了每个收入水平下的 Y 的累积比例 $G_L(r)$ 和人口变量百分比 p。基尼系数的计算公式如下:

$$\mathrm{Gini}(F) = \frac{N/2 - \sum_{1}^{N} G_L(r)}{N/2}$$

其中,N 代表每组的总样本量。在此我们省略了下标“g”。

图 4.1 中的洛伦兹曲线支持两个基尼系数的比较中得到的结论:20 世纪 90 年代早期匈牙利的收入不平等状况比法国更为严重。但是洛伦兹曲线同时显示两个国家中不平等的方式有所不同。与法国相比,匈牙利的不平等程度更高,而且在中高收入水平上更为显著,而在低收入水平上并不太显著。基尼系数是不平等程度的量化表达;而洛伦兹曲线使得比较收入分布的样式更为容易。我们将表明,相对分布方法是进行分布比较的另一种更为形象、在直觉上更具吸引力的视角。

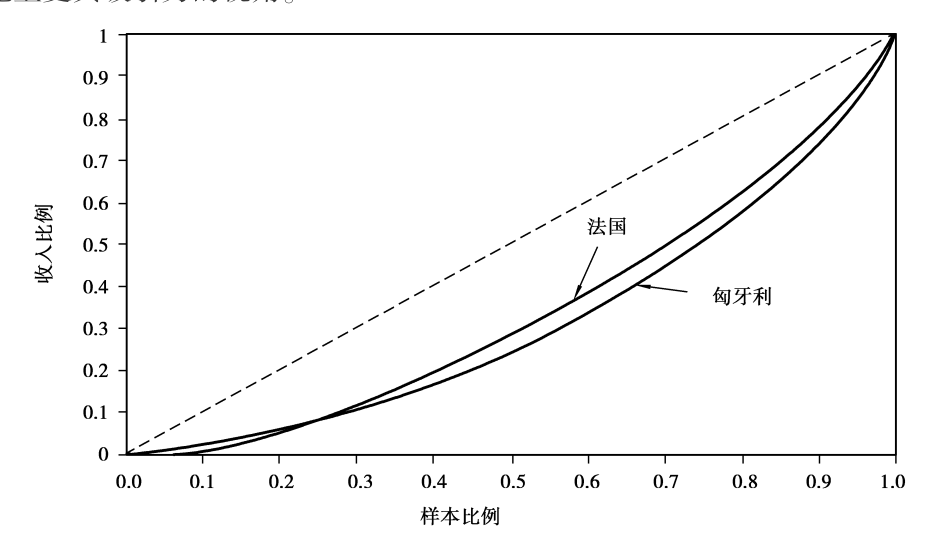



图 4.1　法国(1990)和匈牙利(1992)收入分配的洛伦兹曲线

事实上,正如汉德科克和莫里斯(Handcock and Morris,1998)所言,洛伦兹曲线是基于相对分布的。洛伦兹 PDF 就是通货分布(currency distribution)(可以是美元或英镑、法郎、欧元等)与收入分布的再量纲密度比。所谓通货

分布,就是经济上货币总量分配给每个收入水平的比例。因此,洛伦兹曲线可以视为是相对分布的一种特殊形式。用收入 CDF 对美元分布进行洛伦兹变换,即可得到 $R_L$,而 $G_L(r)$(即洛伦兹曲线)就是相应的相对 CDF。

正如汉德科克和莫里斯(Handcock and Morris,1998)所指出的,洛伦兹曲线与相对分布方法之间存在着三个方面的主要差异:所涉及的组数、测量的单位和量纲不变性(scale invariance)的水平。我们从最后一个方面谈起。如果一个分布是另一个分布乘以一个常数而得来,那么这两个分布的洛伦兹曲线将是完全相同的,在此意义上我们称洛伦兹曲线在乘积方面具有量纲不变性。与此相比,相对分布对于原始测量的所有单调变换都具有量纲不变性。此外,洛伦兹曲线描述的是 Y 的累积比例对应于人口的累积比例,不平等是用洛伦兹曲线对 45°对角线的偏离来测量的;而相对分布画出的对于 Y 的每个水平人口分位数对应于人口分位数。最后,洛伦兹曲线源自于同一组人口的两个特性,因此,在比较两组时就需要两条曲线;而相对分布直接比较两个分布,用一条曲线就可以反映出两个分布间的差异。

我们可用 S-plus 软件中用于相对分布法的程序 reldist.s 来分析收入数据。这一程序是由汉德科克和莫里斯编写的,可以从他们的网址上下载:http://www.stat.Washington.edu/handcock/RelDist/Software/。在那儿也可以得到作为 SAS 宏命令的相对分布法程序。

由于货币兑换率的波动性,加之我们得不到这些收入分配的任何标准转换,更重要的,由于我们的兴趣是比较相对分布和不平等程度。因此,我们把匈牙利收入变量乘以一个常数因子,以使两个中位数有相同的绝对值。

如图 4.2 和图 4.3 所示为比较 1990 年法国和 1992 年匈牙利收入分布的相对分布曲线。图 4.2 代表的是 1992 年匈牙利对 1990 年法国收入分布的相对 CDF。两个国家之间的差异可用水平和垂直的栅格线来描述。我们已将两个中位值设定为相同,因此在中位值处,$r=G(r)$。我们可以看出,匈牙利样本中有 7% 位于法国样本的 10 等分中最低一级,而法国样本中有 12% 位于匈牙利分布中的 10 等分中最低一级。在分布的上端,匈牙利样本中有 84% 位于法国样本的 10 等分中第九级及以下,而法国样本中有 95% 位于匈牙利分布中的 10 等分中第九级及以下,这两个比例是较为相近的。两个分布在第三个、第八个和第九个 10 等分中的分歧最大。假定中位值收入水平相同,匈牙利比法国有更多的人收入低于中位值,但是也有更多的人收入高于中位值,因此,出现了一个波浪纹形状的曲线。这表明匈牙利有着更为严重的不平等。

图 4.2 描述了 1990 年法国和 1992 年匈牙利收入分布的相对 CDF 曲线,而图 4.3 显示了 1992 年匈牙利对 1990 年法国收入分布的相对 PDF。大于 1 的值(即图中的虚线上方的部分)表明比较组(1992 年的匈牙利)比参照组(1990 年的法国)有着更大的密度。而小于 1 的值表示密度更小,在此实际值是作为一个乘积因子存在。看起来,主要的差异在于分布的最底端和最顶端。匈牙利中只有很少的收入部分落入到了法国分布 10 等分中的最低一等分。

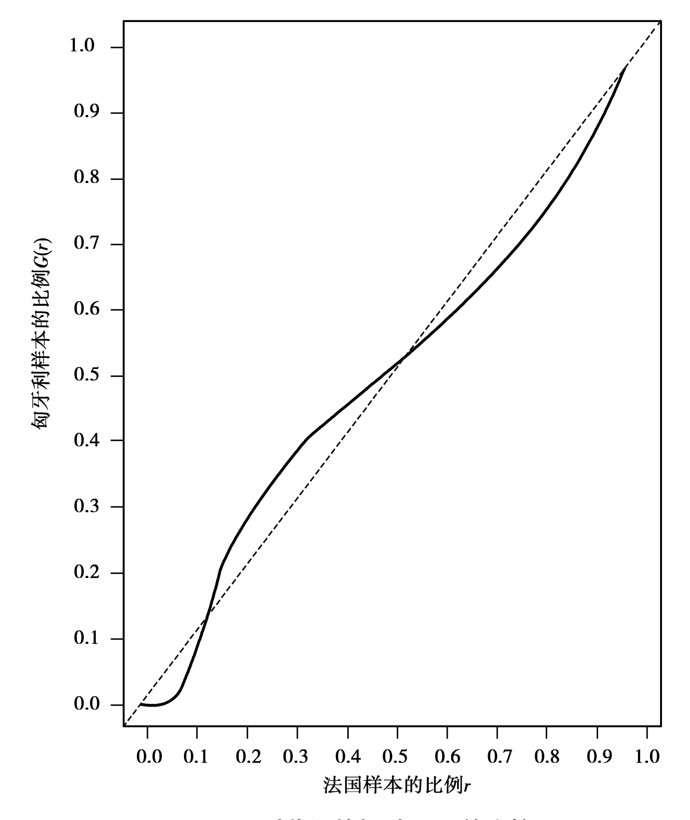



图 4.2 对收入的相对 CDF 的比较

但是,落入法国的第二个 10 等分中的匈牙利收入为法国的 1.2 ~2.8 倍,而落入法国的第三个 10 等分到第八个 10 等分中的匈牙利收入只有法国的 1/2。但是落入法国收入的 82% 以上的匈牙利人逐渐又多了起来,达到了法国的 2.5倍。该图再一次说明了匈牙利的不平等程度高于法国。这种方法以另一种方式更为直观地展现了两个国家在 20 世纪 90 年代收入不平等程度上的差异。

这个例子仅仅展示了研究相对分布的可能方式之一。使用相对分布进行比较可以给出比这更多的信息。更进一步的分析包括对相对分布进行分解以及调整协变量。例如,比较组的分布可能只是参照组分布的简单位移,两个分布之间的差异可以归结为这种简单的位移。在对于位置差异进行调整后,剩余的差异可以解释是形状上的差异——一般包括离散范围、偏度和其他统计性质。相对分布可以分解为位置差异和形状差异。此外,分布差异可能受到协变量分布变化的影响。例如,工资分布的比较可以受到教育分布变化的影响。更为精细的相对分布方法可以将相对分布分解为几个部分,分别代表某些协变量的边缘分布变化的效应。如果想进一步了解分解相对分布的细节,请参阅汉德科克和莫里斯(Handcock and Morris,1998,1999)。

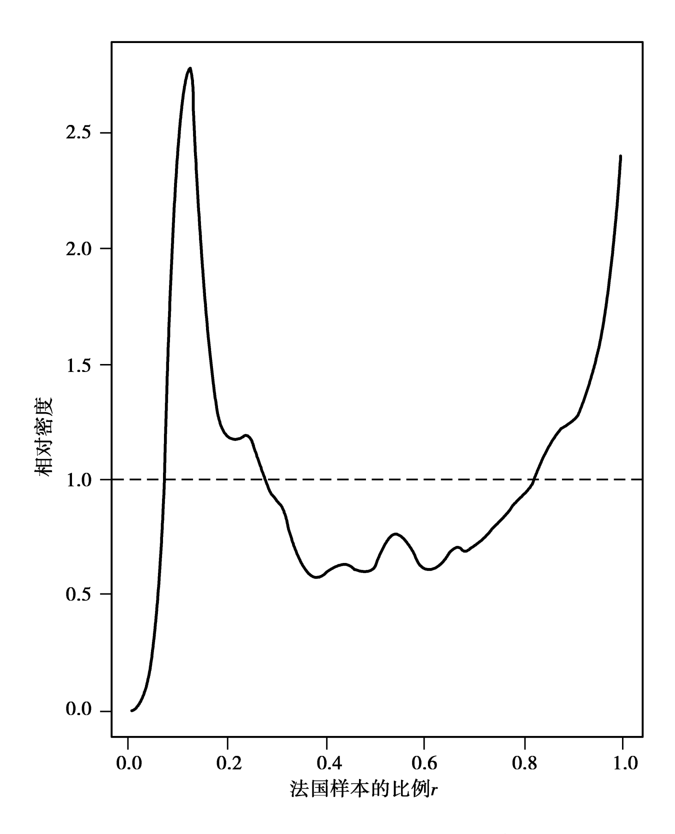



图 4.3　对收入的相对 PDF 的比较

第 5 章

# 比率的比较

Comparison of Rates

比率(rates)的简单比较可以采用第 3 章中描述的标准 t 检验来进行,但有时数据包含一些混杂因素,如果不考虑这些因素,简单比较就是无效的。例如,在人口学和生物医学中,死亡率和发病率经常被用作比较。众所周知,这些比率在不同年龄、种族、地理区域中的分布并不一致。在进行横截面或者纵贯性比较时,这些成分因素的差异将会导致比率的差异,即使纯比率(或者说各种特定成分中的比率)根本没有差异。尽管比率可以在一个包含大量混杂因素并能以控制的线性模型中作为结果被分析,如同前面章节中的处理方法一样,但是研究者通常还是宁愿将处理比率的方法作为一种专项论题。在本章中,我们将考察两种被广泛使用的进行比率的比较的方法——标准化和分解。

## 5.1 数 据

首先让我们来考察一下美国在两个时期(1970 年和 1985 年)的死亡率数据。众所周知,在 15 年间美国的死亡率是下降了,但相应的美国的人口成分也发生了变化。表 5.1 中的数据给出了死亡数和按年龄和种族分类的人口数量。在 1970 年和 1985 年美国居民的粗死亡率分别为每 1 000 人中 9.422 和 8.739。粗略地看,两个比率非常接近,表明美国死亡率在 15 年间并无明显的下降。但是,使问题复杂化的是美国变动着的年龄结构和种族成分,数十年来美国人口变得越来越老化,非白种人的比重越来越大。我们必须调整死亡率以便将这种变动着的人口结构考虑进去。

表 5.1　美国 1970 年和 1985 年的死亡率

| 年份 | 年龄 | 白人 | | 非白人 | |
|---|---|---|---|---|---|
| | | 死亡数 | 人口数 | 死亡数 | 人口数 |
| 1970 | 0 | 54 876 | 2 968 | 19 791 | 535 |
| | 1 ~ 4 | 8 624 | 11 484 | 2 924 | 2 162 |
| | 5 ~ 14 | 13 537 | 34 614 | 3 310 | 6 120 |
| | 15 ~ 24 | 35 510 | 30 992 | 9 751 | 4 781 |
| | 25 ~ 34 | 28 286 | 21 983 | 10 907 | 3 096 |
| | 35 ~ 44 | 54 282 | 20 314 | 18 335 | 2 718 |
| | 45 ~ 54 | 138 871 | 20 928 | 30 640 | 2 363 |
| | 55 ~ 64 | 265 133 | 16 897 | 43 240 | 1 767 |
| | 65 ~ 74 | 393 721 | 11 339 | 51 810 | 1 149 |
| | 75 ~ 84 | 456 247 | 5 720 | 33 556 | 448 |
| | 85 + | 232 540 | 1 315 | 14 415 | 117 |
| 1985 | 0 | 27 864 | 3 041 | 12 166 | 707 |
| | 1 ~ 4 | 5 351 | 11 577 | 1 988 | 2 692 |
| | 5 ~ 14 | 6 812 | 27 450 | 2 121 | 6 473 |
| | 15 ~ 24 | 30 390 | 32 711 | 7 545 | 6 741 |
| | 25 ~ 34 | 38 463 | 35 480 | 13 389 | 6 547 |
| | 35 ~ 44 | 49 608 | 27 411 | 16 207 | 4 352 |
| | 45 ~ 54 | 92 204 | 19 555 | 24 430 | 3 034 |
| | 55 ~ 64 | 241 237 | 19 795 | 45 243 | 2 540 |
| | 65 ~ 74 | 422 956 | 15 254 | 59 690 | 1 749 |
| | 75 ~ 84 | 513 954 | 8 022 | 54 894 | 804 |
| | 85 + | 389 513 | 2 472 | 29 538 | 236 |

注：人口数的单位为千。

资料来源：Liao（1989a，表 1）。

## 5.2 标准化

进行调整的一种常用方法被称为**标准化**，它使用一个标准组或标准总体的统计信息来计算某一组或某一总体中的事件比率。标准可以是另一总体，或是前期的同一总体，或是两个（或多个）总体的数学平均。如果统计信息是成分，这种方法被称为**直接**标准化；而各种特定成分中的比率（composition-specific rates）被用作进行调整的标准时，这种方法被称为**间接**标准化。详尽的讨论可参阅弗莱舍（Fleiss，1981）。

### 5.2.1 直接标准化

对表 5.1 中的两个数据序列进行粗死亡率比较，可以得出每 1 000 人中死亡率的变化为 −0.683（=8.739 − 9.422）。死亡率在 15 年间是否仅仅下降了这么一点儿呢？为了回答这个问题，我们可以假定有一个标准的人口结构，然后计算新的死亡率来进行比较。通常的一个选择是对各成分所占比例进行数学平均。在这个例子中，使用前一时期人口的成分比例作为计算后一时期人口的死亡率的标准是很合理的。让我们将表 5.1 中的数据看做是从列联表中得到的数据，然后考虑表中单元格（i，j，k）中的观察频数。下标"i"表示人口成分中的某一特定类别；下标"j"表示要比较的某一特定组或特定总体；下标"k"表示结果变量中的某一特定事件或者状态（在当前例子中是指活着还是死亡）。在某一时点某一年龄和种族组中活着的人数就是人口总数减去死亡数。对于组 j，事件 k 在特定成分中的比率的定义如下：

$$r_{ij(k)} = \frac{f_{ijk}}{f_{ij+}} \tag{5.1}$$

其中，下标"+"表示边缘总计，即沿被代替的下标进行的加总。第 j 组或第 j 个总体中事件的粗比率为（the crude event rate）：

$$r_{j(k)} = \frac{f_{+jk}}{f_{+j+}} \tag{5.2}$$

其中，$f_{+j+}$ 是第 j 组中的个案总数。在当前例子中，我们只对 k=1 感兴趣，表 5.1 中报告的处于风险中的人口总数已经是沿 k=1，2 进行的加总。让我们进一步定义如下的成分比例（composition proportion）：

$$p_{ij} = \frac{f_{ij}}{f_{+j}} \tag{5.3}$$

其中，k 被省略掉了，因为成分比例并不涉及事件类别。如果像本例一样只有两组，标准比例（standard proportion）$p_i^s$ 可通过计算 $p_{i1}$ 和 $p_{i2}$ 的算术平均来得到，也可用 $p_i^1$ 或 $p_i^2$ 作为 $p_i^s$。直接标准化后的比率的定义为：

$$r^{ds}_{.j(k)} = \sum_{i=1}^{I} r_{ij(k)} p^{s}_{ij} \qquad (5.4)$$

其中,加总是沿着成分水平 i 的所有 I 个值进行的。应用式(5.4)于 $f_{ijk}$ 的表中,将 $p^{1}_{i}$ 作为标准,我们得到:

$$0.009\ 163 \times 0.001\ 456\ 3 + 0.000\ 462 \times 0.056\ 347 + \cdots + 0.125\ 161 \times 0.000\ 547 = 0.007\ 324$$

即如果 1985 年的人口结构与 1970 年保持一致,每 1 000 人中死亡率将是 7.324,而非 8.739。使用了早期的人口比例作为标准以后,死亡率的变化将会是 -2.098,而非比较粗死亡率时得出的 -0.683。有时,计算相对风险(relative risks)是有用的。在此直接标准化的相对风险为 0.777(=7.324/9.422),这表明在对于年龄和种族成分差异进行了修正之后,美国人口的死亡风险在 1985 年比在 1970 年更低了。

## 5.2.2　间接标准化

另一种方法是间接标准化。假定我们完全不知道后一时期人口中各种特定成分的死亡率,所了解的只是人口的成分比例和粗死亡率(或者事件的发生总数,在本例中是死亡数)。在这样的情况下,我们可采用间接标准化,它假定后一时期的成分结构恰如观察所示,而各种特定成分中的比率则与标准人口相同[①]。

我们使用在直接标准化中的符号,首先假定各种特定成分中的比率与标准人口相同,从而得到死亡期望数:

$$f^{is}_{.j(k)} = \sum_{i=1}^{I} f_{ij+} r^{s}_{ij(k)} \qquad (5.5)$$

即作为计算的标准。将式(5.5)应用于美国死亡率数据中,采用 1970 年人口中各种特定成分的死亡率作为标准,我们得到:

$$3\ 041\ 000 \times 0.018\ 489 + 11\ 577\ 000 \times 0.000\ 751 + \cdots + 236\ 000 \times 0.123\ 205 = 2\ 648\ 193$$

这样,假定年龄和种族结构与 1985 年所观察的相同,但是使用 1970 年人口中各种特定成分的死亡率作为标准,那么死亡的期望数将为 2 678 193。在死亡率或发病率研究中,间接标准化的结果有时被表达为标准死亡率比值(或者标准发病率比值),即所观察的死亡数与各种特定成分中的比率和标准人口相同时的期望死亡数之比值。如果标准人口就是要比较的某一组或某一总体,标准死亡率比就成为一种相对风险(relative risk)。在例中,标准死亡率比值为 0.788(=2 085 563/2 648 193),表明如果 1985 年各年龄、各种族的死亡率与 1970 年相同时,我们将期望还会有略多于 20% 的死亡出现。

间接标准化的比率可以通过粗死亡率与标准死亡率比值相乘而得到:

---

① 即与前一时期的人口相同。——译者注

$$0.788 \times 8.739 = 6.883$$

这一间接调整过的死亡率表明，死亡率变化为 -2.539（=6.883 - 9.422）。最后，比较1985年和1970年美国人口，死亡的间接标准化的相对风险为6.883/8.739 = 0.788。相对风险的倒数为1/0.788 = 1.270，给出了标准死亡率比值，它表明假定1985年各年龄各种族的死亡率与1970年一样，得到的死亡率水平将比1985年的观察值高出27%。

可以看到，直接标准化方法和间接标准化方法给出的结果是不同的。事实上，不同的标准化程序给出不同的标准化比率和相对风险是很正常的。问题的关键是要以某种选定的标准为依据，调整某一组或某一总体的汇总比率（summary rates），以使比较成为可能。这种的标准化便于我们比较多个组的汇总比率，但是它并不能取代各种特定成分中的比率，对总体中各种特定成分中的比率的比较，其含义更为丰富。

## 5.2.3 基于模型的标准化

本节内容假定读者已经了解对数线性模型的基本知识。不熟悉这一问题的读者，首先应当学习对数线性模型。下一章是有关广义线性模型的，将会为理解它提供更多的材料。

我们使用的符号与5.2.1节和5.2.2节是一致的。使用基于模型的频数而非观察频数，我们来考虑事件k在组或总体j中在特定成分中的比率：

$$r_{ij(k)} = \frac{F_{ijk}}{F_{ij+}} \tag{5.6}$$

其中，$F_{ijk}$代表在统计模型中列联表的期望频数。第j组（或总体）的未加调整的事件比率为：

$$r_{j(k)} = \frac{F_{+jk}}{F_{+j+}} \tag{5.7}$$

在此，我们有维度为 $I \times J \times K$ 的 $C \times G \times D$ 三向表，其中，I为用i代表的成分的水平数目；J为用j代表的组（G）的数目，而K为k代表的依变量（D）的类别数目。特定成分的比率取决于对C，G，D交互分类的三向表中的$F_{ijk}$。对这样的表建模，以评估C，G，D之间的关联，这就是清洗方法（Purging method）的理论依据（Clogg，1978；Clogg and Eliason，1988；Clogg，Shockey，and Eliason，1990）。对三向表的一般性乘积模型（或对数线性模型）为：

$$F_{ijk} = \tau\tau_i^C\tau_j^G\tau_k^D\tau_{ij}^{CG}\tau_{ik}^{CD}\tau_{jk}^{GD}\tau_{ijk}^{CGD} \tag{5.8}$$

两因素的τ参数表示三个因素中的某两个因素的偏关联。其中，最重要的是参数$\tau^{CG}$和$\tau^{CGD}$，它们测量了由于成分与分组存在相关而造成的混杂效应，以及分组与依变量之间的相关取决于成分的水平的程度。

如果我们不考虑与D的相关，而是考虑$C \times G$的二向表，上述模型将变为：

$$F_{ij+} = \gamma\gamma_i^C\gamma_j^G\gamma_{ij}^{CG} \tag{5.9}$$

在此,参数 $\gamma^{CG}$ 表示成分与分组之间的交互关系,被称为**边缘**的(marginal) CG 交互关系。而 $\tau^{CG}$ 被称为**偏**(partial) CG 交互关系。

清洗方法用上述两个模型估计一组调整后(或清洗后的)的比率,它使用以下公式:

$$r^{*}_{ij(k)} = \frac{F^{*}_{ijk}}{F^{*}_{ij+}} \tag{5.10}$$

其中,频数 $F^{*}$ 通过用 $\tau^{CG}$ 和 $\tau^{CGD}$,或者 $\gamma^{CG}$(具体要由建模者来决定)来调整(分解)相应的频数 F 而得到。调整后的总比率是清洗方法的目标,可以用下式计算:

$$r^{*}_{\cdot j(k)} = \frac{F^{*}_{+jk}}{F^{*}_{+j+}} \tag{5.11}$$

其中,$r^{*}_{\cdot j(k)}$ 表示 j组中每一组的调整过的总比率。

依据所选择的用于调整的参数不同,四种清洗方法如下定义:

偏 CG 调整:

$$F^{*}_{ijk} = \frac{F_{ijk}}{\tau^{CG}_{ij}} \tag{5.12}$$

偏 CG 和 CGD 调整:

$$F^{*}_{ijk} = \frac{F_{ijk}}{\tau^{CG}_{ij}\tau^{CGD}_{ijk}} \tag{5.13}$$

边缘 CG 调整:

$$F^{*}_{ijk} = \frac{F_{ijk}}{\gamma^{CG}_{ij}} \tag{5.14}$$

边缘 CG 和三因素 CGD 调整:

$$F^{*}_{ijk} = \frac{F_{ijk}}{\gamma^{CG}_{ij}\tau^{CGD}_{ijk}} \tag{5.15}$$

上述公式中的清洗产生了一组调整后的频数,它们去除了 C 与 G 相关的影响,或者去除了 C,G 和 D 之间相关的影响。现在我们将这些方法应用于美国死亡率数据中。表 5.2 显示了以 1970 年人口为标准的 1985 年的标准化比率,使用了全部四种清洗方法以及直接和间接标准化的结果。

清洗后的比率用 Purge 程序计算得出,Purge 程序可在斯科特· 伊莱亚森(Scott Eliason)的类别数据分析系统(CDAS)中运行。它可以在如下网址得到:http://www.soc.umn.edu/~eliason/CDAS.htm。

似乎存在着一个规律,即边缘 CG 调整给出的估计值与直接标准化方法最接近。在当前的例中,小数点后三位完全没有区别。偏 CG 调整给出的结果与间接标准化方法较为接近。偏 CG 和 CGD、边缘 CG 和三因素 CGD 通常比其他方法对数据表的调整程度更大。如前所述,在选择标准化方法方面没有标准;一种不同的标准化方法和一个不同的标准人口将导致一组不同的结果。

表 5.2 使用各种标准化方法后的美国死亡率

| 方法 | 死亡率 | | 差异 |
|---|---|---|---|
| | 1985 | 1970 | |
| 粗死亡率 | 8.739 | 9.422 | -0.683 |
| 直接标准化 | 7.324 | 9.422 | -2.098 |
| 间接标准化 | 6.883 | 9.422 | -2.539 |
| 偏 CG 调整 | 6.939 | 9.422 | -2.483 |
| 偏 CG 和 CGD 调整 | 6.438 | 9.422 | -2.984 |
| 边缘 CG 调整 | 7.324 | 9.422 | -2.098 |
| 边缘 CG 和 CGD 调整 | 5.618 | 9.422 | -3.904 |

注:各种方法使用的调整标准都是 1970 年的美国人口。
计算清洗方法所使用的软件为:Eliason 的 CDAS。

## 5.3 分 解

只要比率已经被标准化,两个比率就可以比较。但是,有时我们可能会想搞清楚,比率的差异或变化中有多少可归因于成分的差异或变化,又在多少可归因于实际的比率差异或变化。分解方法在这方面更进了一步,它可以把比率的差异分解为各种成分的差异和比率自身的差异。

在北川(Kitagawa,1955)将标准化方法扩展到分解方法的经典研究中,她说明了在粗比率差异中存在最多三个混杂因素的影响时如何进行分解,她的两因素四部分方法(two-factor four-component method)受到了相当重视。随后的许多发展都是对她的分解法的进一步精细化。有关这些方法的完整评论可以参见廖(Liao,1989a)和达斯• 格普塔(Das Gupta,1991)。随后的文献也对于北川的方法提出了进一步的精细化。大多数方法都依赖于某种数学调整,但廖(Liao,1989a)提出了一种基于模型的方法。我们在下面除了考察北川的经典方法外还将考察这种基于模型的分解方法。

### 5.3.1 算术分解法

北川(Kitagawa,1955)的两因素分析可以在有两个因素影响比率时对两个比率进行分解。因素 A 有 L 个类别,用 l 表示;因素 B 有 M 个类别,用 m 表示。如前所述,我们将有一个维度为 $L \times M \times J \times K$ 的 $A \times B \times G \times D$ 四向表,其中,J 是组(G)的数目,以 j 表示,K 是依变量(D)的类别数,以 k 表示。使用前一节中对观察频数的符号,我们将北川的四个部分:比率效应(RE),A 因素效

应(AE),B 因素效应(BE),因素 A 和 B 的联合效应(ABE)分别定义为:

$$RE = \sum_{l}\sum_{m} \frac{\frac{f_{lm1+}}{f_{++1+}} + \frac{f_{lm2+}}{f_{++2+}}}{2} (r_{lm1(1)} - r_{lm2(1)}) \tag{5.16}$$

$$AE = \sum_{l}\sum_{m} \frac{r_{lm1(1)} + r_{lm2(1)}}{2} \frac{\frac{f_{+m1+}}{f_{++1+}} + \frac{f_{+m2+}}{f_{++2+}}}{2} \left( \frac{f_{lm1+}}{f_{+m1+}} - \frac{f_{lm2+}}{f_{+m2+}} \right) \tag{5.17}$$

$$BE = \sum_{l}\sum_{m} \frac{r_{lm1(1)} + r_{lm2(1)}}{2} \frac{\frac{f_{l+1+}}{f_{++1+}} + \frac{f_{l+2+}}{f_{++2+}}}{2} \left( \frac{f_{lm1+}}{f_{l+1+}} - \frac{f_{lm2+}}{f_{l+2+}} \right) \tag{5.18}$$

$$ABE = \sum_{l}\sum_{m} \frac{r_{lm1(1)} + r_{lm2(1)}}{2} \times \frac{\frac{f_{lm2+}}{f_{+m2+}}\frac{f_{+m1+}}{f_{++1+}} - \frac{f_{lm1+}}{f_{+m1+}}\frac{f_{+m2+}}{f_{++2+}} + \frac{f_{lm2+}}{f_{l+2+}}\frac{f_{l+1+}}{f_{++1+}} - \frac{f_{lm1+}}{f_{l+1+}}\frac{f_{l+2+}}{f_{++2+}}}{2} \tag{5.19}$$

由于我们现在要比较的只有两组(两个人口群体),关心的事件只有一件(死亡),因此,在式(5.19)中下标“j”和“k”处只用 1 或 2 即可。应用此公式于美国死亡率数据中,我们得到了年龄效应(AE)为 1.556,种族效应(BE)为 0.061,年龄和种族的联合效应(ABE)为 −0.081,比率效应(RE)为 −2.228。这些成分效应的总和等于粗比率差异 −0.683。

## 5.3.2　基于模型的分解法

上面分析的四向表可被视为一种 C × G × D 三向表,因为 C 总是可被视为包括了 A 和 B 的多个因素的交叉分类。在这样的三向表中,最后一个维度是我们想了解的事件。前一节考察过的清洗方法可以被用来分解此类三向表中的比率差异。为了达到这一目标,廖(Liao,1989a)将清洗方法扩展为基于模型的分解方法。

为了估计基于模型的成分效应,需要进行以下三个步骤:

1. 建立联立线性方程系统如下:

$$\begin{aligned} RE + AE + BE + ABE &= CRD \\ RE + \quad\quad BE + ABE &= RD(A) \\ RE + AE + \quad\quad ABE &= RD(B) \\ RE \quad\quad\quad\quad\quad\quad &= RD(AB) \end{aligned} \tag{5.20}$$

其中,RE,AE,BE 和 ABE 定义如前。CRD 为粗比率差异。RD(A),RD(B),RD(AB)代表在因素 A、因素 B、因素 A 和 B 的影响分别被清洗之后的比率差异。方程系统之所以这样建立,目的在于要使左端的每一项都可通过一个方程减另一个方程而消掉其他项,最终得到解。

2. 在系统中我们需要基于四个已知数求解四个未知数,四个已知数中有

一个是观察到的,其他三个 RD(A),RD(B),RD(AB)是通过清洗方法估计出来的。因此,这一步骤中还包括了估计。

3. 最后,我们将估计出的 RD(A),RD(B),RD(AB)代入方程组,得到未知数的解。

步骤1 和3 是相当直截了当的。让我们主要关注步骤2,说明一下清洗比率差异的估计方法。为了用 CDAS 估计出 RD(A),RD(B),RD(AB),必须对用于清洗方法的输入数据进行相应整理,以使 C 反映出相对应的因素 A,B 以及 AB,其目的在于利用三向的 $C \times G \times D$ 设置。

为了估计出 RD(A)和 RD(B),要先在排除掉的那个变量的每个水平上调整要研究的因素的比率,然后再根据那个排除掉的变量的观察频数分布来按比例地汇总。例如,如果对于 A 的比率要被清洗掉混杂效应,那么,首先要在维度 B 的每一个水平上用清洗方法评估三向的 $A \times G \times D$ 表,然后再用 B 的频数分布的观察比例来加权,以合并为一个调整后的汇总比率(summary rate)。由于因素 A 而导致的比率差异由下式给出:

$$RD(A) = r_{.j(k)}^{A*} - r_{.J(k)}^{A*} \qquad 其中\ j \neq J \tag{5.21}$$

其中,J 是其汇总比率与其他组的汇总比率进行比较的那个组。最简单的情形(即 $J=2$ 时)只有一个比较。调整后的比率 $r_{.j(k)}^{A*}$ 定义如下:

$$r_{.j(k)}^{A*} = \sum_{m} r_{m\,j(k)}^{A*} \frac{f_{m\,j+}}{f_{+\,j+}} \tag{5.22}$$

其中,上标"A"表示比率中已经清洗掉 A 的影响。因为清洗方法是对 M 个三向的 $A \times G \times D$ 分表分别进行的,调整后的比率必须再用权数 $f_{m\,j+}/f_{+\,j+}$ 合并为一个调整后的汇总比率。这样就给出了清洗掉 A 的影响后的比率。类似地,通过把 A 换为 B,把 m 换为 l,即可得到了清洗掉 B 的混杂影响后的比率 $r_{.j(k)}^{B*}$。

RD(AB)是清洗掉 A 和 B 的联合影响后的组间比率差异。输入 $A \times B$ 作为 C 中的成分,这样就可直接得到 RD(AB)。所得的比率可写为 $r_{.j(k)}^{AB*}$。估计完这些值之后,则可计算 AE,BE,ABE 和 RE 的值。由艾伯特• 切范(Albert Chevan)编写的一个计算机程序可以执行这种方法,它可以在以下网址得到:http://staff.uiuc.edu/~f-liao/decomposition/。

下面,我们将基于模型的分解方法应用于美国死亡率数据中。在标准化中,自然需要将某一组作为参照组,但在进行分解时却不要求这样,这是因为我们感兴趣的是两个比率的**绝对**差异的组成部分。相比之下,标准化的关注点是相对于作为参照组的某一标准来调整比率,因此,使用某一特定组作为参照更为可取。故表 5.3 中报告的分解结果并不依赖于一个标准组。

表 5.3　美国 1970 年和 1985 年的死亡率差异的效应分解

| 方法 | 粗死亡率差异 | 组成部分 | | | |
|---|---|---|---|---|---|
| | | 年龄 | 种族 | 联合 | 残余死亡率差异 |
| 北川 | -0.683 | 1.566 | 0.061 | -0.081 | -2.228 |
| 编 CG | -0.683 | 2.015 | 0.040 | -0.050 | -2.687 |
| 偏 CG 和 CGD | -0.683 | 2.436 | 0.017 | 0.329 | -3.464 |
| 边缘 CG | -0.683 | 1.568 | 0.060 | -0.080 | -2.231 |
| 边缘 CG 和 CGD | -0.683 | 1.983 | 0.037 | 0.297 | -2.999 |

注:由于存在四舍五入的误差,组成部分的加总并不一定等于粗死亡率。
计算各组成部分的效应所用的软件为切范(chevan)编写的 decompos。

使用基于模型的方法而得到的各组成部分的效应应近似于从更为传统的分解方法中得到的结果。例如,边缘 CG 清洗方法给出的结果非常接近于北川的方法。没有标准组的边缘 CG 清洗方法只处理成分和分组之间的交互作用,这一点上它很接近直接标准化方法(Clogg and Eliason,1988)。与此类似,北川的方法是通过平均不同组中 C 的比例来处理 C×G 的交互作用。表 5.3 中的结果支持这种观点。结果还显示,在基于清洗的数种方法之间存在着很大的变异。最强烈的对比存在于考虑三因素交互作用和不考虑三因素交互作用的方法之间。这表明 CG 关联与 D 的水平相关,它是一个混杂因素。注意,在三因素影响被清洗后,AB 的联合效应是正的。这一发现在直觉上看很有趣,因为它暗示年龄和种族除了各自的效应之外,它们共同作用构成了联合效应掩盖了真正的死亡率差异。

那么研究者应当采用哪一种方法呢? 正如廖(Liao,1989a)建议的,如果我们视问题为偶发问题(a contingency problem),其中,维度 C 在调整之后就可压缩掉;那么,“偏清洗”就是合适的方法。但是,如果我们视 C 和 G 为解释变量,它们通过清洗可正交化,那么,就应当考虑使用“边缘清洗”方法(也可参阅 Clogg,Shockey,and Eliason,1990)。此外,如果我们认为 CGD 交互作用应当被清洗,那么,就应当使用包含有 CGD 交互作用的某一种模型方法。如果有人只是想把成分的差异平均化,那么,使用北川经典的组成部分方法就足够了。

基于模型的分解方法可以很方便地应用于有着不止两个成分因素和不止两个组的数据。如果想进一步了解这种方法,请参阅廖的文章(Liao,1989a)。总而言之,基于模型的方法比传统的平均化方法更为灵活,特别是在涉及多个混杂因素和多个组时。切范的软件 decompos 支持对于多个因素和多个组的比率分解。

第6章

# 广义线性模型中的比较

Comparison in Generalized Linear Models

## 6.1 导 论

本章所关注的是被称为广义线性模型(以下简称 GLM)的统计模型中的分组比较。应用研究者用到的许多模型都是 GLM 的例子,包括线性回归、对数单位模型(logit)、概率单位模型(Probit)、对数线性模型(loglinear)以及泊松回归模型等。麦卡拉和纳尔德(McCullagh and Nelder,1989)对这一问题进行了详尽阐释,多布森(Dobson,1990)给出了非常好的导论。GLM 的一般形式如下:

$$E(Y) = \mu \tag{6.1}$$

$$\eta = g(\mu) \tag{6.2}$$

$$\eta = X\beta \tag{6.3}$$

其中,Y 是一个独立同分布的随机变量,其概率分布属于指数族。解释变量 X、参数向量 β 构成一个线性预测值 η,而这一线性预测值 η 又通过某一联系函数 g(·)与 y 的期望向量 μ 相关联。

我们还可用另一种方式来概括 GLM,这样就要使用不同于式(6.1)~式(6.3)的设定。GLM 由以下部分组成:

1. 随机部分:Y 这一部分是独立同分布的,属于指数族,由各种特定分布(distribution-specific)的参数来描述 E(Y)和方差。例如,如果 Y 服从正态分布,则有 E(Y) = μ,有固定方差 $\sigma^2$。

2. 系统部分:向量 X 的元素 $x_1, x_2, \cdots, x_k$ 的组合构成了一个线性预测值 η,即:

$$\eta = \sum_{k=1}^{K} \beta_k x_k$$

系统部分对于 GLM 的所有成员都完全相同。

3. 随机部分和系统部分之间的联系,以一个单调可微的联系函数 g(·)来设定:在具有正态分布(或高斯分布)的线性回归模型中,联系函数是恒等

函数，即：

$$\eta = \mu$$

其他联系函数将有不同的形式。

各种 GLM 的第 2 部分是不变的，因此，我们在下面的 6.1.1 节和 6.1.2 节中将考察各种 GLM 的第 1 部分和第 3 部分的各种变化。

## 6.1.1　指数分布族

令 y 为 Y 的实现值。y 的每个元素服从一个指数族中的分布，形式如下：

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right) \tag{6.4}$$

其中，$a(\cdot)$，$b(\cdot)$，$c(\cdot)$ 是一些函数，随指数族中的分布不同而不同（Barndorff-Nielsen，1978，1980；McCullagh and Nelder，1989；Nelder and Wedderburn，1972）。如果离散参数 $\phi$ 已知，分布就是具有"典范参数（canonical parameter）"$\theta$ 的指数分布。如果离散参数未知，分布既可能是，也可能不是两参数的指数族。$\mu$ 和 $b(\theta)$ 之间的关系是 $\mu = b'(\theta)$。通过定义式（6.4）右端的各组成部分，我们则可定义指数族中的各种不同分布。

例如，当 $\theta = \mu$，$\phi = \sigma^2$，$a(\phi) = \phi$，$b(\theta) = \theta^2/2$，$c(y,\varphi) = -\frac{1}{2}[y^2/\sigma^2 + \ln(2\pi\sigma^2)]$时，指数族形式就成为正态分布：

$$f(y;\theta,\phi) = \exp\left[\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

这表明了正态密度分布与指数族之间的关系。

在泊松分布中，$\theta = \ln\mu$，$\phi = 1$，$a(\phi) = \phi$，$b(\theta) = \exp(\theta)$，$c(y,\phi) = -\ln y!$，于是：

$$f(y;\theta,\phi) = \exp\{(y\ln\mu - \mu) - \ln y!\}$$

$$= \frac{e^{-\mu}\mu^y}{y!} \qquad 其中\ y = 0,1,2,\cdots,\infty$$

其中，$\mu$ 就是通常的泊松参数。具有离散随机变量 Y 的泊松分布函数可以被视为指数族的一个特例。

让我们再来以二项分布为例来进行说明。令 $\theta = \ln[\mu/1-\mu]$，$\phi = 1$，$a(\phi) = \phi$，$b(\theta) = n\ln(1+e^\theta)$，$c(y,\phi) = \ln\binom{n}{y}$，指数族形式则变为二项分布：

$$f(y;\theta,\phi) = \exp\left\{y\ln\mu - y\ln(1-\mu) + n\ln(1-\mu) + \ln\binom{n}{y}\right\}$$

$$= \binom{n}{y}\mu^y(1-\mu)^{n-y} \qquad 其中\ y = 0,1,2,\cdots,\infty$$

其中，y 是在 n 次独立试验中观察到随机变量 Y 成功的次数，而随机变量 Y 的成功概率或成功均值为 μ，它在每次试验中都相同。这表明二项分布是指数族的成员之一。在此没有讨论的分布还有指数分布、极值分布（冈贝尔分布）、伽玛分布、反高斯分布、负二项分布及帕累托分布。

当式（6.4）中的 φ 是已知常数时，上式则简化成为自然指数族的公式：

$$f(y;\theta) = a(\theta)b(y)\exp\{yd(\theta)\} \tag{6.5}$$

通过下面的式子：$a(\theta) = \exp[-b(\theta)/a(\phi)]$，$b(y) = \exp[c(y,\phi)]$，$d(\theta) = \theta/a(\phi)$，我们可把式（6.5）与式（6.4）联系起来。包含多余参数（nuisance parameter）φ 的两参数分布，如正态分布和伽玛分布，则可用式（6.4）来描述；而单参数分布，如二项分布和泊松分布，则可用式（6.5）更方便地描述。

### 6.1.2 联系函数

联系函数定义了 Y 的期望值和基于一组自变量的线性预测值 η 之间的关系。联系函数区分了 GLM 模型族中的各种类型，而联系函数可有许多类型，因此，我们将主要关注的是在 GLM 的常见类型中使用到的联系函数，首先简要地介绍耐受性分布（tolerance distribution）的概念[①]。

在毒理学中，为了分析某一化学剂量对于受试者的致命性，经常要应用二分结果模型（binary outcome models）。令 x 为某一化学物质的剂量，Y 为二分结果（生/死）。令 T 代表受试者对于剂量的耐受性水平；那么，T 可被视为一个隐于其后的不可观察的连续随机变量，因此，T-Xβ 的分布形式则不同于 Y 的分布形式。例如，结果数据（或者正态单位模型的随机部分）服从二项分布，但这一模型的耐受性分布却服从正态分布。在这种意义上讲，联系函数也可被视为对于二分结果模型的两种不同分布的联系。当对这一特点有所了解后，我们可对常用的联系函数进行考察。

1. 恒等：

$$\eta = \mu$$

恒等联系函数定义的是基于正态分布的传统线性回归模型。

2. 对数：

$$\eta = \ln \mu$$

对数联系用于基于泊松分布的泊松回归模型。

3. 对数单位：

$$\eta = \ln\left(\frac{\mu}{1-\mu}\right)$$

使用这一联系函数，我们设定的是一个基于二项分布的且有着二元响应

① "耐受性"在医学上指某种药物剂量或刺激强度不会引发生物体不良反应的范围。——译者注

变量的对数单位模型。这一模型的耐受性分布服从逻辑斯蒂分布。

4. 概率单位:

$$\eta = \Phi^{-1}(\mu)$$

其中,$\Phi^{-1}$是标准正态累积分布函数的反函数。这一联系函数用于基于二项分布的且有着二分响应变量的概率单位模型中。这一模型的耐受性分布是正态的。

5. 多项对数单位:

$$\eta_j = \ln\left(\frac{\mu_j}{1-\mu_J}\right)$$

其中,j代表响应类别1,…,J中的第j个类别。这一联系函数是对数单位联系函数的自然扩展,用于基于多项分布的且有着多项结果变量的多项对数单位模型。实际上,二项分布也可视为多项分布的一个特例。

6. 互补双对数:

$$\eta = \ln[-\ln(1-\mu)]$$

这一联系函数也用于分析二项数据,但是它的耐受性分布服从极值分布(the extreme value distribution)。

对于6.1.1中讨论的指数族的每个成员,都存在一种特殊的联系函数,被称为典范联系函数(canonical link):

$$\theta = \eta$$

其中,θ是式(6.4)中定义的典范参数。例如,三种常见分布的典范联系函数如下:

1. 正态分布:$\eta=\mu$;

2. 泊松分布:$\eta=\ln\mu$;

3. 二项分布:$\eta=\ln[\mu/(1-\mu)]$。

在含有典范联系函数的GLM中估计β时,存在简单的充分统计量$X'y$。也就是说,$X'y$中包括了所有关于β的信息,因而足以用于估计β。

## 6.1.3　最大似然估计

为了简便起见,令$f(y;\theta)$代表下面的联合概率密度函数:

$$f(y_1,y_2,\cdots,y_N;\theta_1,\theta_2,\cdots,\theta_K)$$

上述联合概率密度函数是基于N个观察和K个参数。由于Y的各项都是彼此独立的,联合概率密度函数也可表达为:

$$f(y;\theta) = \prod_{i=1}^{N} f(y_i;\theta_1,\cdots,\theta_K) \tag{6.6}$$

从数学表达式上看,似然函数$L(\theta;y)$与联合概率密度函数$f(y;\theta)$完全相同。但是,$f(y;\theta)$强调的是在含有固定参数θ的参数模型给定时随机变量Y的概率分布;而$L(\theta;y)$表达的是在数据y给定时如何选择或估计参数θ。

令Θ为包括了参数向量θ的所有可能取值的参数空间。θ的最大似然估

计值 $\hat{\theta}$ 是使似然函数最大化的估计量，而使似然函数最大化等价于使对数似然函数$\pounds(\theta;y) = \ln L(\theta;y)$最大化，而对数似然函数更易于处理。上述可用数学符号表示为：

$$\pounds(\hat{\theta};y) \geqslant \pounds(\theta;y) \qquad \text{for all } \theta \text{ in } \Theta \tag{6.7}$$

具体到式(6.4)中，其对数似然函数为：

$$\pounds(\theta,\phi;y) = \sum_{i=1}^{N} \frac{y_i\theta - b(\theta)}{a(\phi)} + \sum_{i=1}^{N} c(y_i,\phi) \tag{6.8}$$

我们最终的目的是获得模型参数 $\theta$ 的一般形式。具体来说，回归类型的估计量 $\hat{\beta}$ 可通过求解似然方程得到，即将对数似然函数的一阶(偏)导数设定为零：

$$\frac{\partial\pounds(\beta,\phi;y)}{\partial\beta_k} = 0$$

令$\partial\pounds = \partial\pounds(\beta,\phi;y)$，则：

$$\frac{\partial\pounds_i}{\partial\beta_k} = \frac{\partial\pounds_i}{\partial\theta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_k}$$

得到的解能否保证对数似然函数处于最大值，这可通过检验而得到确证，即当 $\beta_k = \hat{\beta}_k$ 时对数似然函数的二阶导数是否为负。

$$\frac{\partial^2\pounds(\beta,\phi;y)}{\partial\beta_k\partial\beta_j} < 0$$

在 GLM 中的最大似然估计可以使用迭代重加权最小二乘法。$\hat{\beta}$ 向 $\beta$ 的收敛率取决于信息矩阵，信息矩阵有元素 $E[\partial^2\pounds(\beta,\phi;y)/\partial\beta_k\partial\beta_j]$，等于如下表达式：

$$I = X'WX$$

其中，$W$ 是对角矩阵，主对角元素为：

$$w_i = \frac{(\partial\mu_i/\partial\eta_i)^2}{Var(y_i)}$$

信息矩阵也用于估计 $\beta$ 的方差与协方差。

## 6.2 比较广义线性模型

### 6.2.1 零假设

试考察以下两个分别针对两组观察而建立的 GLM，观察总数为 N(为了便于阐述，观察分为两组，分别为 $i = 1,\cdots,M$ 和 $i = M+1,\cdots,N$)：

$$\eta_1 = X_1\beta_1 \tag{6.9}$$

$$\eta_2 = X_2\beta_2 \tag{6.10}$$

其中，$X_1$ 和 $X_2$ 分别是维度为 $M \times K$ 和 $(N-M) \times K$ 的解释变量矩阵，$\beta_1$和 $\beta_2$都是 $K \times 1$ 的参数向量，$\eta_1$和 $\eta_2$分别是维度为 $M \times 1$ 和 $(N-M) \times 1$ 的两组的线性预测值。与第 3 章中均值相同的零假设相似，检验两个 GLM 相同的零假设为：

$$H_0: \quad \beta_1 = \beta_2 \tag{6.11}$$

我们可以把对两个组的比较扩展为对 G 组的比较，则零假设变为：

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_G \tag{6.12}$$

两组比较中的备择假设为两组的 GLM 参数向量不相同，而多组比较中的备择假设为多组的 GLM 参数向量存在显著不同。

### 6.2.2　使用似然比检验进行比较

我们要检验式（6.11）中的零假设，则应使用似然比检验（以下简称 LRT）。在两组比较的情形中，LRT 统计量为：

$$LRT = -2(\pounds_R - \pounds_U)$$

其中，约束模型的对数似然为：

$$\pounds_R = \sum_{i=1}^{N} \ln L_i(\hat{\beta}, \phi; y_i) = \pounds(\hat{\beta}) \tag{6.13}$$

而非约束模型的对数似然为：

$$\pounds_U = \sum_{i=1}^{M} \ln L_{1i}(\hat{\beta}_1, \phi_1; y_{1i}) + \sum_{i=M+1}^{N} \ln L_{2i}(\hat{\beta}_2, \phi_2; y_{2i})$$
$$= \pounds(\hat{\beta}_1) + \pounds(\hat{\beta}_2) \tag{6.14}$$

要检验式（6.11）中的假设，则应计算$\pounds_R$和$\pounds_U$之间的差距：

$$LRT = -2\pounds(\hat{\beta}) - \{[-2\pounds(\hat{\beta}_1)] + [-2\pounds(\hat{\beta}_2)]\} \quad \sim \chi^2$$

因此，要检验两组的 GLM 参数组之间的相同性，只要从对 $H_0$（即约束模型）估计的 $-2$ 倍的对数似然函数中减去从 $H_1$（即非约束模型）估计的 $-2$ 倍的对数似然函数之和。也就是说，先分别对两组观察进行 GLM，把得到的两个 $-2$ 倍的对数似然函数相加，然后再从汇总模型（pooled model）中得到的 $-2$倍的对数似然函数统计量中减去这个和。得到的 LRT 统计量渐进服从卡方分布，自由度等于两个模型中参数数目之差。将上述程序扩展到对多组的参数相同性（即式（6.12）中的假设）检验中，则：

$$LRT = -2\pounds(\hat{\beta}) - \sum_{g=1}^{G} [-2\pounds(\hat{\beta}_g)] \quad \sim \chi^2 \tag{6.15}$$

事实上，式（6.15）中的检验就是对 LRT 的直接应用。注意，如同 GLM 中参数的最大似然估计一样，这一检验也是基于大样本这一特性的。

我们也可以采用沃尔德检验（以下简称为 WT）作为另一种检验方法。按式（6.11）进行零假设，在 X 变量固定的假定下，通常的 WT 公式则变为：

$$WT = (\hat{\beta}_1 - \hat{\beta}_2)'[Var(\hat{\beta}_1) + Var(\hat{\beta}_2)]^{-1}(\hat{\beta}_1 - \hat{\beta}_2) \tag{6.16}$$

其中，Var(·)是参数的方差—协方差估计矩阵。WT 与 LRT 是渐进等效的，在许多软件中都可以很容易地得到 $Var(\hat{\beta})$。

另一种渐进等效的检验是拉格朗日乘数检验(以下简称 LMT)。由于它采用的梯度向量(gradient vectors)被称为"得分"，故称为得分检验。它由如下公式给出：

$$LMT = \left(\frac{\partial \pounds_R}{\partial \hat{\beta}}\right)' [I(\hat{\beta})]^{-1} \left(\frac{\partial \pounds_R}{\partial \hat{\beta}}\right) \qquad (6.17)$$

其中，中间一项为信息矩阵。LMT 只依赖于约束模型。由于非约束模型可表示为包括 $\hat{\beta}$ 和分组变量的交互项的汇总模型，加上目前计算机的高速运行，因此进行得分检验的一种简单方法是，对约束模型和非约束模型都进行 LMT 估计，然后取两个得分统计量的差值。

根据式(6.15)采用 LRT 进行多组比较是极其简便的，但是对于 WT 和 LMT 来说远非如此。在本章后面的例中，我们将主要关注两个组之间的参数相同性检验。

### 6.2.3 作为一种特例的邹氏检验

对于线性模型来说，联系函数是恒等的，随机部分服从正态分布，误差项的方差是恒定的；此时著名的邹氏检验可以视为 6.2.2 节中讨论的 LRT 的一个特例。内尔德和韦德伯恩将 LRT 统计量称为"(有量纲的)偏差(scaled deviance)"(Nelder and Wedderburn, 1972)，或 D。当联系函数为恒等、分布为正态时，用于检验两组相同性检验的式(6.14)中的 LRT 则变为：

$$\begin{aligned} D = & \left[\frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \mu_i)^2 - N\ \ln(2\pi\sigma^2)\right] \\ & - \left[\frac{1}{\sigma_1^2}\sum_{i=1}^{M}(y_{1i} - \mu_{1i})^2 - M\ \ln(2\pi\sigma_1^2)\right] \\ & - \left[\frac{1}{\sigma_2^2}\sum_{i=M+1}^{N}(y_{2i} - \mu_{2i})^2 - (N - M)\ln(2\pi\sigma_2^2)\right] \end{aligned} \qquad (6.18)$$

同方差性假定或者离散同质性假定(即 $\sigma_1^2 = \sigma_2^2 = \sigma^2$)排除了误差项的方差在各组之间的任何变异，则式(6.18)可进一步简化为：

$$D = \left[\frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \mu_i)^2\right] - \frac{1}{\sigma^2}\left[\sum_{i=1}^{M}(y_{1i} - \mu_{1i})^2 + \sum_{i=M+1}^{N}(y_{2i} - \mu_{2i})^2\right] \qquad (6.19)$$

现在将式(6.19)中的三个组内平方和分别表示为 $D_1, D_2, D_3$，将它们视为三个单独的"偏差"测量。显而易见，$D_1, D_2, D_3 \sim \chi^2$。既然 $D = D_1 - (D_2 + D_3)$，把它除以 $D_2 + D_3$，然后再以它们各自的自由度来对分子和分母进行加权，我们便得到：

$$F = \frac{[D_1 - (D_2 + D_3)]/(2K - K)}{(D_2 + D_3)/(N - 2K)} \tag{6.20}$$

$$= \frac{\left\{ \sum_{i=1}^{N}(y_i - \mu_i)^2 - \left[ \sum_{i=1}^{M}(y_{1i} - \mu_{1i})^2 + \sum_{i=M+1}^{N}(y_{2i} - \mu_{2i})^2 \right] \right\}/K}{\left[ \sum_{i=1}^{M}(y_{1i} - \mu_{1i})^2 + \sum_{i=M+1}^{N}(y_{2i} - \mu_{2i})^2 \right]/(N - 2K)} \tag{6.21}$$

式(6.20)给出了一个服从 F 分布的精确检验统计量，它与式(3.20)中给出的在同方差性条件下对经典线性回归进行的邹氏检验完全相同。为了把式(6.21)和式(6.15)用 D 联系起来，也为了简化起见，我们采用如下表达式：

$$F = \frac{D/K}{(D_2 + D_3)/(N - 2K)} \tag{6.22}$$

因此，式(6.15)中的 LRT 统计量除以含有恒等联系的单组 GLM 的 $-2\mathcal{L}$之和，再分别用它们各自的自由度进行加权，其结果与离散同质性假定下的邹氏检验相同。

## 6.3　对数单位模型例解

对数单位模型是最常见的含有非线性联系函数的 GLM 模型，本节中我们将使用对数单位模型来展示如何检验两个方程中系数相同的零假设。

### 6.3.1　数据

表 6.1 中的数据是克洛格和肖基(Clogg and Shockey，1988)使用的数据的扩充形式，其中包括一个二分的响应变量，表示被访者在 1980 年中是投票支

**表 6.1　1980 年美国白人的总统选举行为**

| 政治态度 | 男　性 | | 女性 | |
|---|---|---|---|---|
| | 里根 | 卡特或其他人 | 里根 | 卡特或其他人 |
| 1 | 2 | 21 | 5 | 21 |
| 2 | 24 | 94 | 41 | 116 |
| 3 | 58 | 112 | 56 | 122 |
| 4 | 235 | 216 | 307 | 319 |
| 5 | 212 | 94 | 161 | 120 |
| 6 | 197 | 55 | 172 | 69 |
| 7 | 25 | 17 | 26 | 13 |
| 总计 | 753 | 609 | 768 | 780 |

资料来源：1982 年、1983 年、1984 年、1985 年的综合社会调查项目(GSS)。

注：政治态度取值为从“1 = 极端自由主义”到“7 = 极端保守主义”。

持里根还是卡特或者另一位候选人。除了性别外,被访者的政治态度指标是另一个解释变量。现在的这个示例使用的样本包括了那一年中投票的 2 910 名白人男性和女性,但是比他们使用的例子多了一个变量。此外,本例中的调查涉及多个年份,而不只有一年。表 6.1 显示的是例中用到的分组数据。

看起来,在政治态度对投票行为的影响上存在着性别差异。但是只有采用 GLM 模型并认真地进行数据研究后,我们才能证明这一观点。

## 6.3.2 模型比较

为了使分析更为简单和简约,我们选择如下形式的线性对数单位模型来拟合表 6.1 中的分组数据:

$$\ln\left[\frac{P(y=1)}{1-P(y=1)}\right] = X\beta$$

这一模型抓住了政治态度类别的定序性质,但是将其作为数量型量度来处理。我们感兴趣的零假设为:

$$H_0: \quad \beta_m = \beta_f$$

其中,每个性别组中的 $\beta_g(g=f,m)$ 都包括两个参数。

**表 6.2 选举数据的对数单位模型的对数似然估计值**

| x-Variable | 1. 单模型检验 | 2. 汇总模型 | 3. 只有女性 | 4. 只有男性 |
|---|---|---|---|---|
| 截距 | -2.347 | -2.153 | -1.963 | -2.347 |
| | (0.217) | (0.144) | (0.193) | (0.217) |
| 是否女性 | 0.384 | — | — | — |
| | (0.290) | — | — | — |
| 政治态度 | 0.595 | 0.527 | 0.463 | 0.595 |
| | (0.049) | (0.033) | (0.044) | (0.049) |
| 是否女性×政治态度 | -0.132 | — | — | — |
| | (0.066) | — | — | — |
| 模型的 $\chi^2$ | 308.783 | 299.596 | 23.061 | 176.366 |
| 模型的自由度 | 3 | 1 | 1 | 1 |
| -2 ln L | 3 719.344 | 3 728.531 | 2 022.830 | 1 696.514 |
| LRT | | 9.187 | | |
| WT | | 9.138 | | |
| LMT | | 7.656 | | |

注:标准误在括号内给出。

我们估计了四个模型(参见表 6.2)。模型 1 包括虚拟变量"是否女性"和其他 x 变量"政治态度"的所有的主效应和交互效应,它代表了检验组间差异的传统虚拟变量方法。模型 2,3,4 都包括有 x 变量"政治态度",我们要检验

的正是在两个性别分组之间它的效应是否相同。模型 2 代表 $H_0$——男性和女性中效应相同的假设,而模型 3 和模型 4 联合代表 $H_1$,分别估计在女性和男性之间的效应。

在此,得出两点评论。首先,模型 4 中的估计值与模型 1 中的主效应估计值完全相同,这是因为模型 4 仅仅包括男性被访者,这等同于在模型 1 中将所有涉及“是否女性”的变量设定为 0。事实上,把模型 1 中“是否女性”和“是否女性 × 政治态度”的效应估计值与模型 4 中的估计值相加,则可得到模型 3 中的估计值。其次,模型 2 给出的一组估计值是模型 3 和模型 4 估计值的“平均”:模型 2 中的估计值总是落在模型 3 和模型 4 估计值之间。为了用图形来比较男性和女性的政治选择,我们在图 6.1 中展示了不同性别投票支持里根的预测概率和观察概率。看起来模型 1 非常好地拟合了数据。性别之间明显存在差异,这可从两条曲线的交叉上看出来。

可方便地计算出 LRT,以评估参数是否相同。利用式(6.15),则 3 728.531 − (2 202.830 + 1 696.514) = 9.187。用模型 3 和模型 4 中估计参数个数的总和减去模型 4 中估计参数个数,可得到自由度为 2。统计量在 0.025水平下是显著的,在 0.01 水平下几乎是显著的。注意,使用模型的 $\chi^2$ 值也可得到同样的结果。同时,LRT 和传统的联合交互效应检验(即模型 1)给出了相同的结果;某一种方法是否优于另一种方法,要依情形而定。例如,只关心某一些系数是否相同的研究者可能会偏好有着交互项的虚拟变量方法。但是,如果不只存在两组,研究者也不仅仅关心几个 x 变量在各组间的异同,那么,虚拟变量的处理方法就会变得更复杂。

利用式(6.16),我们得到了 WT 统计量为 9.138,这一结果与似然比统计量几乎完全相同。通过使用模型 2 和把两个涉及性别变量的参数都设定为零(即使用模型 1 和模型 2 的得分统计量),我们得到了拉格朗日乘数统计量为 7.656,这一结果在 0.025 水平也是显著的。因此,这三个渐进等效的检验给了我们相同的结论——拒斥不同性别之间参数相同的零假设。

## 6.4　风险率模型例解

### 6.4.1　模型

风险率模型这种统计分析类型在不同学科中有着不同的名称:在社会科学中称为“事件史分析”;在工程学中被称为“故障时间分析”;在人口学和医学中称为“生存分析”或“持续期限分析”。读者可参考任何一本教科书来了解此模型的基本知识(如 Allison, 1984; Blossfeld, Hamerle, and Mayer, 1988; Blossfeld and Rohwer, 1995; Courgeau and Leliévre, 1992; Cox and Oakes, 1984; Kalbfleisch and Prentice, 1980; Lancaster, 1980; Lawless, 1982; Tuma and Hannan,

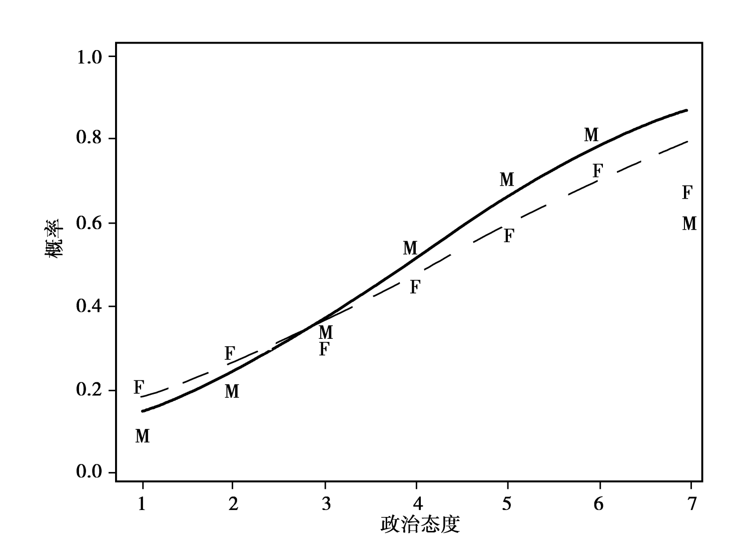



图 6.1 不同性别投票支持里根的预测概率和观察概率

1984;Vermunt,1997;Yamaguchi,1991)。本节的目的只是在 GLM 模型内简要介绍此模型,以便进行比较分析。

风险率模型中的几个核心概念使它有别于其他统计模型。首先,依变量的值取决于观察的相对**状态**。每一个观察在某一时点上都呈现出某种可能的状态。**事件**是从一种状态到另一种状态的转变。例如,某人初次结婚,他就从未婚状态转变成了初婚状态。**风险集**是指在某一时点上可能经历特定事件的所有个人的集合,**风险期**是指某人处于特定事件风险中的时间段。事件史分析也被称为持续期限分析,其原因就在于它分析了在风险期内事件不发生的持续期限。

另外两个重要的概念就是时变性协变量(或依变量)和截删。**时变性协变量**是指那些随时间变化的自变量。例如,年龄和就业状况会随时间变化而改变,而种族背景和性别则不随时间变化而改变。截删可以分为两种:左截删和右截删。为了解释在某一特定事件不发生的持续期限方面的个体差异,我们不仅需要了解影响这一过程的那些协变量,而且还需要了解个体进入(或开始)风险集($\tau_b$)的具体时间,以及事件发生时的具体时间($\tau_e$)。事件不发生的持续期限 T 就是 $\tau_e - \tau_b$。但是,我们可能得不到某些研究对象的 $\tau_b$ 或(和)$\tau_e$ 的信息。如果 $\tau_b$ 没有被观察到,此时发生的就是左截删;如果 $\tau_e$ 没有被观察到,此时发生的就是右截删。风险率模型可以比较容易地处理右截删数据;左截删的处理则更为复杂(除非 $\tau_b$ 已知)。

令 T 为连续的非负随机变量,代表事件不发生的持续期限;f(t) 为 T 的概率密度,F(t) 为 T 的累积分布函数。f(t) 与 F(t) 之间的关系为:

$$f(t) = \lim_{\delta t \to 0} \frac{P(t \leqslant T < t + \delta t)}{\delta t} = \frac{\partial F(t)}{\partial t}$$

另一方面:

$$F(t) = P(T \leqslant t) = \int_0^t f(s)d(s)$$

其中,$\delta t$ 代表非常小的时间区间。生存概率或生存函数是直到 $t$ 时事件不发生的概率,定义如下:

$$S(t) = 1 - F(t) = P(T \geqslant t) = \int_t^{\infty} f(s)d(s)$$

风险率或者风险函数 $h(t)$ 测量的是即时风险,$h(t)\delta t$ 近似等于事件还未发生的条件下在时间 $t$ 和 $t+\delta t$ 之间发生的条件概率。风险率定义如下:

$$h(t) = \lim_{\delta t \to 0} \frac{P(t \leqslant T < t + \delta t \mid T \geqslant t)}{\delta t} = \frac{f(t)}{S(t)} \tag{6.23}$$

风险率模型的类型取决于风险率的分布,更准确地说是取决于对风险率分布的处理方法。

#### 6.4.1.1　参数风险模型

这些模型假定 $T$ 与 $h(t)$ 之间的关系为某一特定函数形式。对于基本函数 $h(t)$,$f(t)$ 或 $F(t)$ 来说,可能的函数形式有很多种(如指数分布、分段指数分布、冈珀茨分布、威布尔分布、对数逻辑斯蒂分布、对数正态分布、伽玛分布和反高斯分布)。

用 GLM 的术语讲,$\eta = X\beta$;如果模型中存在一些时变性协变量,则 $\eta = X(t)\beta$。现在,令 $h(t \mid X)$ 为一个有着协变量 $X$ 的观察在 $T = t$ 时的风险函数。指数模型是最简单的参数风险率模型。它假定生存函数为指数形式(即风险率不随时间变化):

$$h(t \mid X) = \exp(\eta) \tag{6.24}$$

这个模型中 $h(t)$ 只取决于 $X$。相比之下,冈珀茨模型进一步扩展了指数模型,允许 $h(t)$ 依赖于时间的单调变换:

$$h(t \mid X) = \exp(\eta + \alpha t) \tag{6.25}$$

在此模型中 $\ln h(t)$ 被假定为是 $T$ 的线性函数。更进一步扩展就成了威布尔模型,它假定 $\ln h(t)$ 是 $T$ 的对数函数。

$$h(t \mid X) = \exp(\eta)\alpha t^{\alpha-1} \tag{6.26}$$

其中,$\alpha > 0$。

至此描述的所有参数模型都是**成比例的**或者是**对数线性的**。之所以称它们是成比例的,是因为 $X$ 对 $h(t)$ 的效应是相乘的,不存在 $X$ 与时间的交互作用;之所以称它们是对数线性的,是因为 $h(t)$ 的对数是协变量和时间效应的线性函数。有一些参数风险模型既不是成比例的,也不是对数线性的。对数逻辑斯蒂模型就是这样一个示例:

$$h(t \mid X) = \frac{\alpha[\exp(\eta)]^{\alpha} t^{\alpha-1}}{1 + [\exp(\eta)t]^{\alpha}} \tag{6.27}$$

其中，$\alpha>0$，对数逻辑斯蒂模型可以用于非单调的风险率。

#### 6.4.1.2 考克斯的半参数风险模型

参数模型要求了解 T 的分布。考克斯的半参数风险率模型也被称为考克斯比例风险模型，它在应用研究中使用非常普遍，因为它对 T 的分布形式不做任何假定。模型中包括一个未设定的 T 的函数，形式为任意的基准风险函数 $h_0(t)$：

$$h(t|X)=h_0(t)\exp(\eta) \tag{6.28}$$

模型把 X 和 h(t)之间的关系设定为对数线性关系，就此而言，它是参数性的。这个看似简单的模型估计起来却较困难，因为需要在未设定 $h_0(t)$ 的条件下估计 β。考克斯(Cox，1972，1975)提出了部分似然估计法，有效地解决了这一问题。

#### 6.4.1.3 离散时间模型

有时，事件发生时间是不太精确地用几个有限的时间段来测量的。在这种情况下，我们就要用到离散时间模型，它对事件在第 r 个时间段内发生的条件概率进行回归，条件是事件在此时间段之前还没有发生。我们用 $h(t_r)$ 来代表这一概率。注意，$h(t_r)$ 只是对 h(t) 的近似。

存在几种离散时间模型，其中最常用的模型采用对数单位联系函数，结果为：

$$h(t|X)=\frac{\exp(\alpha_r+\eta)\,t^{\alpha-1}}{1+\exp(\alpha_r+\eta)} \tag{6.29}$$

式(6.29)也可以用对数单位形式重新加以表达：

$$\ln\left[\frac{h(t_r|X)}{1-h(t_r|X)}\right]=\alpha_r+X\beta$$

其中，$\alpha_r$ 是一个关于时间段的二分变量。只要进行一些数据操作准备，通常的对数单位模型就可应用于事件史数据。

### 6.4.2 数据

我们用于风险率模型的数据来源于克瑞尔等人的研究(Krall，Uthoff，and Harley，1975)，他们的分析数据是关于多发性骨髓瘤的研究。在这项研究中，他们用烃化基剂来治疗 65 位病人，其中，48 位在研究时去世，17 位在研究结束时仍然存活。数据在本章附录中给出(见表 6.4)，变量“Time”记录的是从诊断起以月计的生存时间，变量“Status”有两个取值 0 和 1，分别表示病人在研究结束时仍然存活还是已经去世。这样，如果“Status”的取值为 0，“Time”的相应取值就是右截删的。

表 6.4 中仅仅包括九个 x 变量中的两个：“log BUN”表示在诊断时用于检验肾功能的血清尿素氮指标(正常值应低于 log 19)，“hemoglobin”表示在诊断

时的血色素指标(男性正常范围为 14 ~ 18,女性正常范围为 12 ~ 16)。这两个变量是通过“向前似然比选择程序”而选择出来的,只有这两个变量是与生存与否显著相关的因素。

### 6.4.3　模型比较

只要采用的是离散时间模型和参数模型,涉及风险率模型的比较就与 GLM 中的比较遵循同样的规则,因为它们全都是对于 GLM 的直接应用。但是,考克斯的比例风险模型还包括额外的一项 $h_0(t)$,因此,就涉及这一项在各组间是否相同的假设检验问题。出于这种考虑,我们介绍用考克斯回归对数据进行分析。

除了参数相同的假设外:

$$H_0: \quad \beta_g = \beta_h \qquad 其中\ g \neq h$$

我们还对以下假设感兴趣:

$$H_0: \quad h_{0g}(t) = h_{0h}(t) \qquad 其中\ g \neq h$$

我们进行分层分析(stratified analysis),允许每一层有自己的基准风险函数,但每一层中的系数被限定为相同的。大多数执行考克斯比例风险模型的统计软件包都可以运行分层分析。与在其他 GLM 一样,为了检验 β 在各组间是否相同,我们要估计三个模型,一个是汇总模型,另外两个模型分别针对两个组,然后进行 LRT 检验。

我们在表 6.3 中展示了四个考克斯比例风险模型的结果。第一列是包括了所有个案的模型,没有按照年龄变量的两个类别来对基准风险函数进行分层。第二列是对所有观察的分层模型。第三和第四列是对两个年龄组的模型,一个是针对 55 岁及以下的病人,一个是针对 55 岁以上的病人。年龄对于生存与否不是一个显著的影响因素,但是它确实有分层效应。

**表 6.3　对附录中的数据进行考克斯模型的比较**

| x 变量 | 1. 所有个案 | 2. 分层所有个案 | 3. 年龄≤55 | 4. 年龄≥56 |
|---|---|---|---|---|
| Hemoglobin | -0.119 | -0.123 | -0.138 | -0.116 |
| | (0.058) | (0.061) | (0.151) | (0.066) |
| log BUN | 1.674 | 1.945 | 3.441 | 1.170 |
| | (0.612) | (0.634) | (1.148) | (0.787) |
| $-2\ln L, h_0(t)$ | 309.716 | 249.711 | 76.093 | 173.617 |
| $-2\ln L, \exp(\eta)$ | 297.767 | 236.208 | 65.101 | 168.273 |

注:标准误在括号中给出。

$h_{01}(t) = h_{02}(t)$ 的假设是否正确?在前两个模型中两个 β 参数被限定为在两组中相等,而 -2 倍对数似然数函数的第二行代表包括协变量的模型中

的对数似然函数,我们取其前两个数字之差值,结果是(297.767 −236.208) = 61.559。我们可得到从 $h_0(t)$ 的 −2 倍对数似然数函数中(即从没有协变量的模型中)得到类似的差异,但是,因为如果没有协变量,模型是不同的,故其结果不完全相同。

至于第二个假设 $\beta_1 = \beta_2$,我们可利用式(6.15)和 $\beta_g$ 完全相同和不同时的分层的 $h_0(t)$ 的 −2 倍对数似然数函数得出结论,236.208 −(65.101 + 168.273) =2.834(df =2),因此,假设得到证实。比较不同模型的参数估计可知,在两个年龄组中 log BUN 的效应有相当大的差异。那么,这个参数在两组中是否存在显著差异呢?应用式(6.16),我们得到 WT 统计量为 2.662(df = 1),结果是不显著的。这表明,β 的大部分差异确实可以归结为 log BUN 的效应差异。总而言之,两个年龄组有着显著不同的基准风险函数,但是在协变量的效应上差异并不显著。

# 6.A 6.4 节中用到的数据

表 6.4 中收集的是 6.4 节中所讨论的示例中采用的数据。

**表 6.4 用烃化基剂来治疗多发性骨髓瘤的生存状态**

| Time | Status | Log BUN | Hemoglobin | Time | Status | Log BUN | Hemoglobin |
|---|---|---|---|---|---|---|---|
| 1.25 | 1 | 2.22 | 9.4 | 7 | 1 | 1.98 | 9.5 |
| 1.25 | 1 | 1.94 | 12 | 7 | 1 | 1.04 | 5.1 |
| 2 | 1 | 1.52 | 9.8 | 7 | 1 | 1.18 | 11.4 |
| 2 | 1 | 1.75 | 11.3 | 9 | 1 | 1.72 | 8.2 |
| 2 | 1 | 1.3 | 5.1 | 11 | 1 | 1.11 | 14 |
| 3 | 1 | 1.54 | 6.7 | 11 | 1 | 1.23 | 12 |
| 5 | 1 | 2.24 | 10.1 | 11 | 1 | 1.3 | 13.2 |
| 5 | 1 | 1.68 | 6.5 | 11 | 1 | 1.57 | 7.5 |
| 6 | 1 | 1.36 | 9 | 11 | 1 | 1.08 | 9.6 |
| 6 | 1 | 2.11 | 10.2 | 13 | 1 | 0.78 | 5.5 |
| 6 | 1 | 1.11 | 9.7 | 14 | 1 | 1.4 | 14.6 |
| 6 | 1 | 1.42 | 10.4 | 15 | 1 | 1.6 | 10.6 |
| 16 | 1 | 1.34 | 9 | 88 | 1 | 1.18 | 10.6 |
| 16 | 1 | 1.32 | 8.8 | 89 | 1 | 1.32 | 14 |

续表

| Time | Status | Log BUN | Hemoglobin | Time | Status | Log BUN | Hemoglobin |
|---|---|---|---|---|---|---|---|
| 17 | 1 | 1.23 | 10 | 92 | 1 | 1.43 | 11 |
| 17 | 1 | 1.59 | 11.2 | 4 | 0 | 1.95 | 10.2 |
| 18 | 1 | 1.45 | 7.5 | 4 | 0 | 1.92 | 10 |
| 19 | 1 | 1.08 | 14.4 | 7 | 0 | 1.11 | 12.4 |
| 19 | 1 | 1.26 | 7.5 | 7 | 0 | 1.53 | 10.2 |
| 24 | 1 | 1.3 | 14.6 | 8 | 0 | 1.08 | 9.9 |
| 25 | 1 | 1 | 12.4 | 12 | 0 | 1.15 | 11.6 |
| 26 | 1 | 1.23 | 11.2 | 11 | 0 | 1.61 | 14 |
| 32 | 1 | 1.32 | 10.6 | 12 | 0 | 1.4 | 8.8 |
| 35 | 1 | 1.11 | 7 | 13 | 0 | 1.66 | 4.9 |
| 37 | 1 | 1.6 | 11 | 16 | 0 | 1.15 | 13 |
| 41 | 1 | 1 | 10.2 | 19 | 0 | 1.32 | 13 |
| 41 | 1 | 1.15 | 5 | 19 | 0 | 1.32 | 10.8 |
| 51 | 1 | 1.57 | 7.7 | 28 | 0 | 1.23 | 7.3 |
| 52 | 1 | 1 | 10.1 | 41 | 0 | 1.76 | 12.8 |
| 54 | 1 | 1.26 | 9 | 53 | 0 | 1.11 | 12 |
| 58 | 1 | 1.2 | 12.1 | 57 | 0 | 1.26 | 12.5 |
| 66 | 1 | 1.45 | 6.6 | 77 | 0 | 1.08 | 14 |
| 67 | 1 | 1.32 | 12.8 | | | | |

资料来源：Krall, Uthoff, and Harley(1975)。

# 第7章 在广义线性模型中进行比较的其他问题

Additional Topics of Comparison in Generalized Linear Models

## 7.1 导 论

在本章中,我们将考察在 GLM 模型中进行比较的其他三个问题:对离散异质性(dispersion heterogeneity)的处理、广义线性的贝叶斯模型及配对研究(matched studies)模型。通常前两个问题可以视为对 GLM 的扩展,但第三个问题只是 GLM 的一种特殊应用。每一个问题都是在新的比较情境中 GLM 的展现。在下面三节中我们将讨论这三个问题。配对研究从定义上讲就是用于比较的,在 7.2 节我们将考察如何用 GLM 处理此类研究。通常的 GLM 假定离散参数 $\phi$ 是个常数,但是一些数据并不满足这一假定。在 7.3 节中,我们将介绍处理离散异质性的办法。GLM 是基于频率论的统计学研究路数的。在 7.4 节中,我们将给 GLM 中的比较注入一些贝叶斯统计思想。

## 7.2 针对配对的个案-对照研究的 GLM

个案-对照设计(case-control designs)广泛地应用于流行病学和医学研究中,研究可以是回顾性的,也可以是前瞻性的。近几年来,这种方法也被应用于社会学研究中,以研究一些稀少事件(Lacy,1997)。史密斯(Smith,1997)讨论了在社会科学的观察研究中用多个对照组配对来估计处理效应。因为在许多社会科学中真正的实验是难于进行的,配对的个案-对照设计可以作为一种控制混杂性因素的行之有效的替代方法。

首先,让我们来区分回顾性研究与前瞻性研究。有两个标准:样本是如何选择的,信息是如何收集的。

- 在回顾性研究中,我们根据目前的结果来抽样,然后探索以前的解释变量取值(也可能包括以前的结果变量取值)。许多个案-对照研究是回顾性的。

- 在前瞻性研究中,我们根据特定解释变量的选择标准来抽样,然后观察发生了什么样的结果。临床试验(clinical trials)和队列研究(cohort studies)常常是前瞻性的。

个案-对照方法与队列方法有关联,但是在许多重要方面亦有不同。队列研究的基础是所有研究对象是否暴露于某因素的记录。然后对暴露者和非暴露者以相同方式进行历时跟踪研究。取决于数据是事件发生之前还是发生之后收集,队列研究可以是回顾性的或者前瞻性的。与此不同,个案-对照研究的基础是一些研究对象展现出某些结果的记录(即个案)。比较组由没有展现出这种结果的个体组成(即对照),然后按照与个案的历史相同的方式记录这些对照者的历史。这样一种设计是回顾性的,但是个案-对照研究也可以是前瞻性的。

潜在的混杂因素在队列方法和个案-对照方法中都很常见。这就是为什么个案-对照研究常常要按照特定的混杂因素来进行配对的原因。非配对的个案-对照数据可以用针对二分变量的 GLM 进行分析,这并不复杂。因此,在这里我们主要关注配对的个案-对照数据的统计分析。

在个案-对照研究中,通常采用三种设计方法:1∶1,1∶m,n∶m 配对。1∶1的配对个案-对照研究中每一个"对子"(pair)中包括一个个案和一个对照;可以把"对子"的概念视为与生存分析中的"风险集合"的概念相对应。1∶m 设计在每一集合中有一个个案和 m 个对照,而 n∶m 设计在每一集合中有 n 个个案和 m 个对照。在配对的个案-对照研究中,是个案还是对照的状况被处理为一种结果,我们感兴趣的是比较个案组与对照组,以发现观察是在处理组(暴露于某因素之下)还是在未处理组(未暴露于某因素之下)。检验处理效应的传统方法是用一种 2×2 设计来看事件的比率或者几率。麦克尼马尔检验(McNemar,1947)就是采用这样一种传统的检验方法。由于还存在除了"处理"之外的其他解释变量,故大多数研究者常常使用条件逻辑斯蒂回归(conditional logistic regression)。

在配对的个案-对照研究中,需要为每对个案-对照集合引入一个新参数。这些参数被称为"多余参数",因为研究者对它们并不感兴趣,只是使得估计更为复杂化了。为了从估计中消除这些多余参数,我们将采用**条件似然**,它是经过调整的似然函数。条件逻辑斯蒂回归就是基于条件似然估计的。

在典型的配对个案-对照研究中,我们感兴趣的是检验如下假设:

$$H_0: \quad \beta_1 = \beta_0$$

而模型为:

$$\ln\left[\frac{P(y_1 = 1)}{1 - P(y_1 = 1)}\right] = X_1\beta_1, \ln\left[\frac{P(y_0 = 1)}{1 - P(y_0 = 1)}\right] = X_0\beta_0 \qquad (7.1)$$

其中,$\beta_1$和$\beta_0$分别是 $X_1$ 和 $X_0$ 的参数向量,通常 $X_1$ 和 $X_0$ 在第一列中包括处理组的状态。下标"1"和"0"表示观察是属于个案还是对照。

注意,$X_1$ 和 $X_0$ 可能长度不同。也就是说,虽然它们在 1∶1设计中有着相

同的行数,但是在 1∶m 和 n∶m 设计中 $X_0$ 的行数至少与 $X_1$ 一样多,但通常都要比 $X_1$ 多。在不同配对集合彼此独立和同一集合中两组响应彼此独立的前提下,与任何逻辑斯蒂回归模型一样,我们可以给出一个联合似然函数。在非配对的个案-对照研究中,两个输入矩阵[①]以及向量 $y_1$ 和 $y_0$ 都可以垂直地联系起来。但是这种方法在此并不奏效。在下面我们将说明如何在 1∶1,1∶m,n∶m 的个案-对照研究中应用条件逻辑斯蒂回归。

### 7.2.1 1∶1的配对研究

如果对于每个个案都只有一个对照,我们应用起条件逻辑斯蒂回归来就简单了,因为 $y_1$ 和 $y_0$ 以及 $X_1$ 和 $X_0$ 都具有相同的维度。在此零假设为个案和对照都同等程度地暴露于某一因素之下。换而言之,即假设不存在处理效应。如果 X 中只有一个测量处理—暴露的因素,则可采用麦克尼马尔检验。观察频数落到了一个 2×2 列联表的四个单元格中(见表 7.1)。

表 7.1 从 1∶1配对组中得到的数据

| | 对照 | |
|---|---|---|
| 病例 | 暴露 | 未暴露 |
| 暴露 | $f_{11}$ | $f_{12}$ |
| 未暴露 | $f_{21}$ | $f_{22}$ |

如果零假设为真,那么不一致的"对子"的期望频次应当等于观察频次的平均数,即 $F_{12}=F_{21}=(f_{12}+f_{21})/2$。依据这种逻辑并加以简化,用观察频次表示的一个卡方统计量为:

$$\chi^2=\frac{(|f_{12}-f_{21}|-1)^2}{f_{12}+f_{21}} \tag{7.2}$$

即为麦克尼马尔检验,其自由度为 1。

更为一般化地,如果在 X 中存在不止一个因素时,使用大多数统计软件中的普通逻辑斯蒂回归程序就可以分析此类数据。这种过程包括以下三个步骤:

1. 取 $X_1$ 和 $X_0$ 的差值 $X_d=X_1-X_0$。事实上,这是把关于"个体"的数据转换成关于"对子(pair)"的数据,$X_d$ 的每一列记录了每个解释变量的差异,包括个案与对照之间的处理状态的差异。

2. 令新的结果变量为常数 1。这也等效于取 $y_1$ 与 $y_0$ 的差值,产生一个有着 N 行的元素全为 1 的向量(N 是对子的数目)。即 $y_1-y_0=1$。

3. 拟合一个没有截距项的逻辑斯蒂回归。在进行了第 2 步后,估计响应

① 即 $X_1$ 和 $X_0$ 的矩阵。——译者注

的均值就不必再有截距项了。

上述程序非常简单,可以直接应用。但是,某些软件(如 SPSS10.0)并不能灵活地拟合一个含有常数结果变量的逻辑斯蒂回归。

### 7.2.2　1∶m 的配对研究

如果在配对集合中对于每个个案不只有一个对照,那么,这种设计就是1∶m设计。在1∶m 设计中,每个对子中的 m 可能是不同的,这种情况下如何进行条件逻辑斯蒂回归估计,这一问题我们将在下一节中讨论更为一般化的 n∶m设计时再去考虑。在此,我们主要关注的是 1∶m 设计的一个特例:即在所有集合中对照的数目 m 是恒定的,并且只有一个协变量("是否暴露")。布雷斯洛曾经讨论过这种特例(Breslow,1982)。

通常的逻辑斯蒂回归仍然可应用于这种特例中。但是,程序与 1∶1设计中的程序不同。它包括如下四个步骤:

1. 整理数据,按照每个集合中被暴露于处理的对象数目(t)来整理观察到的集合数(n)。在总数 n 中,个案被暴露的集合数目被记录为变量 e。

2. 生成一个抵消变量 $o = \ln[t/(m+1-t)]$。

3. 生成一个新变量"处理-暴露",$x = 1$。

4. 对于在 n 次试验中出现 e 次事件拟合一个逻辑斯蒂回归模型,模型包括抵消变量 o 和常数 x,但不包括截距项。

在布雷斯洛的数据中,每个个案有四个对照,总共有 80 个集。用表格形式表示的数据为:

| t | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| e | — | 5 | 19 | 10 | 6 | 0 |
| n | 10 | 20 | 27 | 17 | 6 | — |

上表中的抵消变量为 $o = \ln[t/(5-t)]$。第一列中 e 的缺失值导致了十个集合没能用上,此外不存在五次试验(既包括个案也包括对照)中有五次处理的观察。数据可以用标准的逻辑斯蒂回归程序来进行分析,SAS 的程序员在他们的示范程序中演示过这种分析。

### 7.2.3　n∶m 的配对研究

如果每个集合中有一个或多个个案与一个或多个对照相配对时,我们遇到的就是更为一般性的 n∶m 的配对个案-对照研究设计,1∶1和 1∶m 设计都是它的一种特例。为了说明如何分析这种数据,我们先来看一些抽样数据。

#### 7.2.3.1　数据

数据来自于 SAS 的数据样本库中,包括了 174 位妇女的信息,其中,54 位妇女的婴儿出生体重低于正常标准(故称为"个案"),120 位妇女的婴儿出生

体重正常(故称为"对照")。值得考虑的几项风险因素(即协变量)包括:末次月经时的体重(以下简称 WLM)、怀孕期间是否抽烟(以下简称 SDP),是否有高血压(以下简称 HT),是否有子宫紊动(以下简称 UI)。年龄(age)是作为配对变量的混杂因素。除了年龄和 WLM 外,所有变量都是虚拟变量。如果是个案,变量 LBW 的编码为 1;如果是对照,LBW 的编码为 0。首先所有观察按照年龄来升序排列,再按个案-对照状况排列。本章的附录中列出了所用的数据。

#### 7.2.3.2 个案与对照的比较

对于附录 7. A 中的数据,我们感兴趣的检验假设为:

$$H_0: \quad \beta_1 = \beta_0$$

其中,每个 β 向量都包括 WLM,SDP,HT,UI 的参数,同时年龄这一混杂因素被加以控制,它共有 17 个水平。为了对 n:m 配对数据估计一个条件逻辑斯蒂回归模型,除了自己编写计算机子程序或宏命令外,我们还可以采用专门用于此类模型的程序(如在 Stata 中),也可以采用现成的用于考克斯比例风险模型的计算机程序(如 S-plus,SAS,SPSS),或者采用现成的用于广义线性模型的计算机程序(如 Glim)。用于考克斯回归的部分似然估计完全可以用来计算条件逻辑斯蒂回归模型中的条件似然。下面我们将演示如何用考克斯回归程序来估计这样的条件逻辑斯蒂回归模型。

要用考克斯回归程序来估计条件逻辑斯蒂回归模型,需要有四个步骤:

1. 将个案-对照变量重新编码为失效时间变量(failure time variable)或事件时间变量(event time variable),个案编码为 1,对照编码为 2。在本例中,时间 = 2 − LBW。

2. 生成一个新变量"截删",个案编码为 0,对照编码为 1,或者就使用"事件发生时间"变量。最重要的事是要表明正确的截删取值(在本例中截删为 1,时间为 2)。

3. 使用配对变量作为分层变量。你可以直接把它们作为分层变量引入模型设定中。一些软件只允许包括一个分层变量。在这种情况下,你必须把配对变量编码为一个新的风险集变量。通过对原始的分层变量进行交互分类,可以做到这一点。

4. 如果配对是 1:m 设计,设定 ties = Breslow;如果配对是 n:m 设计,设定 ties = discrete。"ties"是指享有相同的事件发生时间取值的观察。在本例中,设定 ties = discrete。

为了把 n:m 配对的条件逻辑斯蒂回归与风险率模型联系起来,上述程序将每个配对集合作为风险集来处理,对照被作为右截删处理。因此,事件时间变量只有一个数值起到作用。个案是事件发生,而对照是事件未发生。在比例风险模型中分层变量对基准风险函数的参数被去掉了,与此相类似,在条件逻辑斯蒂回归模型中配对变量的参数是"多余参数",也被去掉了。

把配对变量“年龄”作为考克斯回归中的分层变量，利用 SAS 对本章附录中的数据拟合条件逻辑斯蒂回归模型，我们得到了如表 7.2 所示的结果。对事件数目和截删值数目的汇总没有在表中显示出来。如果研究者怀疑分析是否被正确执行，那么，这种汇总是有用的。每一层中（每一风险集中）的事件数应当与每个配对集合中的个案数相同；每一层中的截删观察数应当与每个配对集合中的对照数相同；分层的总数应当等于配对集合的总数。如果有任何不一致，研究者都应当回到程序中去，以确保程序依照上述步骤正确地建立起来。

**表 7.2　为附录 7.A 中的数据进行病例与对照比较**

| x 变量 | β | se(β) | 沃尔德 $\chi^2$ |
|---|---|---|---|
| WLM | -0.015 | 0.007 | 4.500 |
| SDP | 0.808 | 0.368 | 4.822 |
| HT | 1.751 | 0.739 | 5.612 |
| UI | 0.883 | 0.480 | 3.383 |
| LRT, df = 4 | 17.961 | | |
| WT, df = 4 | 15.558 | | |
| LMT, df = 4 | 17.315 | | |

表 7.2 中的估计值给出了一些风险因素对低体重婴儿出生的效应在个案组与对照组之间的比较。结果显示，末次月经体重对于生育低体重婴儿的效应在个案组中（即实际生下了低体重婴儿的妇女）显著低于对照组（没有生下低体重婴儿的妇女）。另一方面，怀孕期间抽烟和高血压对于生育低体重婴儿的影响在个案组中显著高于对照组。子宫紊动的影响在个案组和对照组之间没有显示出显著差异。这样，即使个案组经历了事件而对照组没有经历事件，我们仍然可以在个案和对照间进行 β（对事件发生的效应）比较，因为对照组的信息被处理为截删了。在理论上，两组都可能经历事件的发生和不发生。

## 7.3　离散异质性

GLM 的核心部分就是指数分布族。对于式（6.4）中描述的分布来说，离散参数 $\phi$ 被假定是一个常数，它要么可从数据中直接估计出来，要么是 $\mu$ 的简单函数故是多余的。除了第 6 章中讨论的联系函数 $g(\cdot)$ 外，我们还可以使用一个方差函数 v 来进一步定义 y 的方差和均值之间的关系（Breslow，1996：23；McCullagh and Nelder，1989：29）：

$$\mathrm{Var}(y) = \phi w^{-1} v(\mu) \tag{7.3}$$

其中，方差函数 $v(\cdot)$ 依赖于典范参数，即［$v(\cdot) = b''(\theta)$］，故而也依赖于均值。w 被称为“**先验权数**”（prior weight），对于每个观察都有不同值。例如，对

于一个每个观察都是 m 个独立测量的均值的样本来说，我们有 $w=m$。因此，方差的任何变异一定来自于典范参数（或均值）。

现列举两例来加深理解。对于正态分布而言，$\phi=\sigma^2$，因为联系函数是恒等，而 $v(\cdot)=1$。对于二项分布和泊松模型而言，$\phi=1$，因为它们分别有 $v(\mu)=\mu(1-\mu)$ 和 $v(\mu)=\mu$。

关于 $\phi$ 的假设可能被违反，有以下两种情况：

- 对于泊松和二项数据而言（当 $m>1$ 时），离散可能大于 1（有时也可能小于 1）。这被称为过度离散（或者离散不足），通常是由聚类数据（clustered data）产生，如观察来自于同一单位（同胎所生、家庭、班级和学校等）。而模型中又没有包括对聚类的测量。
- 更为一般化地，$\phi$ 可能是一些自变量的函数。问题的一个特例是线性回归中的异方差性。

换而言之，如果在某一单位中某事件发生的概率依赖于集群的其他单位中事件发生的概率，而且这种依赖是正向的，那么，此时就会发生“过度离散”；如果这种依赖是负向的，那么，这时就会发生“离散不足”。在第二种更为一般化的情形中，GLM 中的统计比较将会受到影响，因为 $\phi$ 值可能在不同组中不同，或者依某些自变量取值不同而不同，或者两者情况共存。以上情况我们定义为离散异质性。

但是，我们不应混淆这两者：y 的方差与 X 的相关；e 的方差与 X 的相关。众所周知，在 GLM 中 y 的方差可以是解释变量的函数，根据式（7.3）因为 y 的方差的定义之中就包括了 $\mu$ 的方差；与此不同，在线性回归模型中误差项的同方差性确保了 $\mu$ 的变异并不包含在误差项中。也就是说，如果 y 的方差仅随 $\mu$ 而变化，这并不违背离散异质性假定。如果误差项随 $\mu$ 之外的另一个量而系统变化，如离散参数 $\phi$ 而在不同观察组间变化，这就违背了离散异质性假定了。

例如，在联系函数为恒等且 $\phi=\sigma^2$ 的线性回归中，方差 $\phi$ 不存在任何特定的名义值，因为 $\phi$ 可以取任何正的数值，只要它不是某一解释变量的函数。如果 $\phi$ 是某个 x 变量的函数，就产生了异方差性问题。同理，适用于其他 GLM 模型。

为了纠正过度离散问题（或离散不足问题），首先我们必须得到离散参数 $\phi$ 的估计值。例如，在二项和泊松数据中常常会出现过度离散问题；离散参数应当为 1（如果存在一个二项先验权数变量 m，离散参数则应为 1/m）。麦卡拉和内尔德（McCullagh and Nelder，1989）建议用下式来估计有着典范联系的 GLM 的离散参数：

$$\hat{\phi}=\frac{1}{N-K}\sum_{i=1}^{N}\frac{(y_i-m_i\mu_i)^2}{m_i\mu_i(1-\mu_i)}=\frac{\chi^2}{N-K} \tag{7.4}$$

如果分析中没有先验权数，则 $m_i=1$。可以看出，估计量其实是用相关的自由度适当加权后的残差平方和。换而言之，$\phi$ 的估计量等于皮尔逊偏差

(Pearson deviance)统计量除以自由度 N－K。虽然皮尔逊偏差统计量常常接近于 GLM 的有量纲偏差(scaled deviance),但它们的定义是不同的。后者的定义如下：

$$2[\pounds(y;y)-\pounds(\hat{\mu};y)]$$

其中,$\pounds(\hat{\mu};y)$代表假定某一模型成立时的最大对数似然,用参数 $\mu$ 表达,拟合的是 y 的观察值向量。在一个饱和模型中(即模型中包括的估计参数数目与独立观察数目一样多),$\hat{\mu}=y$,因此有量纲偏差等于零。自由度等于观察数 N 与模型中参数数目 K 的差值。读者可能已经注意到了,有量纲偏差有着与 LRT 相同的形式,是它的一个特例。唯一的区别在于,偏差选择的参照模型为饱和模型,而 LRT 可以选择任何模型作为参照模型。

通常式(7.4)中描述的估计 $\phi$ 的方法都能得到合理的估计(Breslow,1996;Smith and Heitjan,1993)。估计值的计算很简单,因为许多计算机程序都可以产生皮尔逊偏差统计量。史密斯和海特贾(Smith and Heitjan,1993)研究出了针对 GLM 中过度离散的一种得分检验,他们的程序是用 S 函数执行的,可以从卡内基—梅隆大学互联网上的 Statlib 中得到。即使不进行正规检验,也应当经常性地快速检查一下给定自由度下的偏差。

拉奥和斯科特(Rao and Scott,1992,1999)对于过度离散的集群二项和泊松数据(没有协变量或分组)的处理提出了一种简单方法来加权数据。为了调整过度离散的(有协变量)模型,我们用 $Var(\hat{\beta})$ 乘以 GLM 的量纲因子(scale factor)的估计值,它由式(7.4)给出,或者用此估计值对 y 数据进行加权。或者,对于泊松数据的 GLM 来说,可以用负二项模型来解决过度分布问题。为了在分组比较时调整 GLM 中的离散,我们可用 $\phi_g$ 来调整 $Var(\hat{\beta}_g)$。我们总是可以用 $\phi_g$ 来调整每一组(因为用任何接近于名义值的 $\phi_g$ 进行调整不会对标准误产生很大改变),但是我们仍然考虑有一种显著性检验来决定这种调整是否必要。在对两组 g 和 h($g\neq h$)比较时,基于式(7.4),每组中的皮尔逊卡方可以形成一项 F 比：

$$F_{[N_g-K,N_h-K]}=\frac{\chi_g^2/(N_g-K)}{\chi_h^2/(N_h-K)} \tag{7.5}$$

有必要把含有较大离散的那一组的卡方放在分子中。更为一般地,为了检验 G 组中对于离散同质性的显著偏离,我们可以基于皮尔逊偏差测量进行更为一般的 F 检验：

$$F_{[GK,N-GK]}=\frac{\sum_{g=1}^{G}\left|\chi_g^2-(N_g/N)\chi^2\right|/GK}{\chi^2/(N-GK)} \tag{7.6}$$

其中,N 和 $\chi^2$ 都是指整个样本而言。此 F 检验的目的在于检测出离散异质性,不论是过度离散还是离散不足,它能够足够灵活地包容多组比较中的情况。如果结果显示是显著的,则表明组间离散同质性的假定被违背了。从这

种一般化的 F 检验中得出的结论并不一定与式(7.5)中的成对 F 检验的结论完全相同,因为成对 F 检验的结论要取决于哪两个组被选中进行比较。

要记住,式(6.15)中讨论的 LRT 假定离散同质。如果这一假定不成立,LRT 检验就不再是评估参数相同性的有效方法。在异质的条件下,我们必须用离散因子来调整方差-协方差矩阵。调整后的似然函数将不再是真正的似然,而是准似然函数,虽然其性质与表现与似然函数相同(McCullagh and Nelder,1989,第 9 章)。基于准似然的 GLM 是对于 GLM 的直接扩展(Nelder,1998)。为了检验假设(6.11),在离散异质性的条件下用于进行多组比较的检验(6.15)就变成了准似然比检验:

$$QLRT = -2Q\pounds(\hat{\beta}) - \sum_{g=1}^{G}\left[-2Q\pounds(\hat{\beta}_g)\right] \sim \chi^2 \quad (7.7)$$

还有,我们不能用从调整后的(同质)模型中得到的 −2 倍准似然函数和从未调整的合并(异质)模型中得到的 −2 倍似然函数比较。用 $\phi_g$ 估计调整后的汇总模型(pooled model)相应的准似然函数,$Q\pounds(\hat{\beta})$,可以根据其所属组别用 $\{\hat{\phi}_g\}$ 对 $y_{gi}$ 进行加权而得到:

$$y_{gi}^{(w)} = \frac{y_{gi}}{\hat{\phi}_g} \quad (7.8)$$

其中,$y_{gi}^{(w)}$ 是第 g 组中经过加权的 $y_i$。在分组数据中,$y_i$(事件)和 $n_i$(试验)都应当被加权。在个体数据中,对 $y_i$ 加权意味着对它进行转换,因此也改变了它的期望值,但是 $\hat{\beta}_k$ 未受影响。这种加权方法与克洛格和伊莱亚森(Clogg and Eliason,1987)推荐的在对数线性模型中进行加权的方法是一致的。正如抽样权重可以适用于对数线性模型(Clogg and Eliason,1987),在此讨论的方法可以适用于简单随机抽样设计得到的数据,也可以适用于复杂抽样设计得到的数据。式(7.7)中的 QLRT 统计量可以类比于式(6.15)中的 LRT。同样,我们也可以计算出基于准似然的沃尔德统计量和拉格朗日乘数统计量(即得分统计量)。

在方差同质性条件下,$\phi_g = \phi =$ 名义离散(如在泊松模型和对数单位模型中为 1),这时 LRT 和 QLRT(或者调整前和调整后的沃尔德检验和得分检验)将给出完全相同的结论,因为准似然函数的取值与似然函数相同。在离散异质性条件下,基于准似然的三种检验统计量评估的是离散被调整为同质的各组参数相同性的假设。

另一种用 $\phi$ 估计值来调整 $\hat{\beta}$ 方差的方法使用了得分的观察协方差矩阵而不是期望信息来计算渐进方差,这样就得到了所谓的“三明治方差估计量(sandwich variance estimator)”[①]:

① 这一名称源自于公式的形式,$J^{-1}$ 一前一后如同两片面包,夹住了中间部分。——译者注

$$Var(\hat{\beta}) = J^{-1}\left[\sum_{i=1}^{N} w_i^2\left(\frac{y_i - \mu_i}{v(\mu_i)}\right)^2 x_i x'_i\right] J^{-1}$$

其中,J 是不包括量纲因子 $\phi$ 的费舍尔信息矩阵,$\mu_i$ 用拟合值 $\hat{\mu}_i$ 来估计,方括号中的中间因子是得分的观察协方差矩阵(Breslow,1996)。三明治估计量因其表达形式不同和应用背景不同,也被称为怀特估计量(White extimator)或者休伯估计量(Huber estimator),在主要的统计软件中都可得到这一估计量(如在 SAS 和 Stata 的稳健统计允许对集群数据使用这一估计量)。三明治方差估计量常常会低估中等大小的样本的真实变异性,在这种情况下最好使用再抽样技术来估计方差,如自助法或者刀切法(Jackknife)(Breslow,1996)。

第一种方法是用 $\phi_g$ 来进行调整,它假定 $\phi$ 只随组别 g 而变动。虽然我们可以把这种方法推广到用分组变量之外的因子进行调整,但是因为要对变量的所有水平都估计 $\phi$ 然后再用它加权,随之而来的具体操作将变得非常麻烦。第二种方法利用了三明治统计量,应用起来非常灵活,但它的前提假定是离散参数只随影响 $\mu$ 的 X 变量而变动。这一假定有可能正确,但并非永远正确。此外,从实际操作的角度讲,执行三明治统计量的程序可能并不计算准似然函数来进行比较。这使得 QLRT 的应用更为困难,虽然新的 $Var(\hat{\beta})$ 可以很容易地插入到 WT 的计算中去。

为了更为直接和更为灵活地处理离散异质性问题,存在两种相互关联的建模方法——双重广义线性模型(以下简称 DGLM)(Smyth and Verbyla,1996,1999);均值和离散相加模型(以下简称 MADAM)(Rigby and Stasinopoulos,1995,1996)。两种方法有许多的相似之处,但在普适性程度上有所差别。DGLM 要设定两个联系函数,一个联系函数是:

$$g(\mu) = \eta = X\beta \tag{7.9}$$

另一个联系函数是针对离散的:

$$h(\phi) = \xi = Z\lambda \tag{7.10}$$

其中,$h(\cdot)$ 是不同于 $g(\cdot)$ 的,Z 是影响 $\phi$ 的协变量矩阵,$\lambda$ 包括它们的参数。

DGLM 为 $h(\cdot)$ 设定联系函数的方式与 GLM 为 $g(\cdot)$ 设定联系函数的方式相同,因为 DGLM 中均值和离散的联系函数都是参数性的。广义可加模型(以下简称 GAM)是对 GLM 的扩展,它允许联系函数 $g(\cdot)$ 是非参数的,例如各种各样的平滑函数(smoothing function)(Hastie and Tibshirani,1990)。类似地,MADAM 允许 $g(\cdot)$ 和 $h(\cdot)$ 都是非参数的。因此,MADAM 可被视为 GLM,GAM,DGLM 的一般模型。事实上,DGLM 和 MADAM 都是沿着内尔德指出的两个主要方向对于 GLM 的扩展(Nelder,1998)。在本书中,我们主要关注参数性的联系函数 $h(\cdot)$,正如在第 6 章中我们对函数 $g(\cdot)$ 的讨论。

DGLM 假定 y 服从指数离散模型,这是一个在理论上有说服力的假定。通过限制最大似然法(restricted ML)可以得到 DGLM 的估计。给定 $\lambda$ 任何一

个运行值，DGLM 对用 $w_i/\phi_i$ 加权过的 $y_i$ 使用普通的 GLM 来估计 β。同样，给定 β 任何一个运行值，DGLM 对 $d_i$ 使用伽玛 GLM 来估计 λ，这里的 $d_i = d(y_i, \mu_i)$，表示 $y_i$ 与 $\mu_i$ 之间的直接距离（Smyth，1989）。我们称在 λ 固定时估计 β 的 GLM 为**均值子模型**，称在 β 固定时估计 λ 的伽玛 GLM 为**离散子模型**；后者中的离散参数为 2。通过在 λ 固定时估计 β 和在 β 固定时估计 λ 之间交替变换，这种估计方法非常有效，因为 β 和 λ 都是正交参数（Smyth，1989）。用 S 函数执行的 DGLM 可以在以下网址得到：http://www.maths.uq.edu.au/~gks/s/dglm.html；用 Glim 宏运行的 MADAM 可以在如下网址得到：http://tara.unl.ac.uk/~11stasionpou/madam.html。

在下面的示例中，我们将演示如何用 $\phi_g$ 进行调整，以及如何使用 DGLM。

## 7.3.1 数据

表 6.1 中介绍的数据实际上是对九个地理区域或集群进行汇总后得到的。在此例中，引起离散异质性的潜在集群变量可能是被访者居住的选区。

**表 7.3 1980 年美国白人的总统选举行为**

| 地区 | 性别 | 政治态度 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 新英格兰 | F | 1/3 | 1/7 | 6/18 | 23/45 | 9/16 | 8/12 | 1/1 |
| | M | 1/1 | 0/9 | 3/14 | 16/32 | 12/21 | 10/14 | 1/1 |
| 中大西洋 | F | 0/5 | 9/31 | 7/22 | 64/119 | 17/30 | 18/31 | 2/4 |
| | M | 1/4 | 1/16 | 9/22 | 34/63 | 35/50 | 24/29 | 4/7 |
| 中部偏东北 | F | 1/7 | 7/20 | 9/39 | 50/106 | 37/62 | 35/45 | 4/8 |
| | M | 0/8 | 4/26 | 18/46 | 63/109 | 42/68 | 31/42 | 6/9 |
| 中部偏西北 | F | 1/2 | 4/21 | 5/14 | 34/68 | 22/40 | 9/15 | 1/2 |
| | M | 0/4 | 5/12 | 2/12 | 20/44 | 20/33 | 19/22 | 1/2 |
| 南大西洋 | F | 1/5 | 7/25 | 8/24 | 42/92 | 24/48 | 27/39 | 2/5 |
| | M | 0/3 | 4/21 | 7/18 | 34/64 | 28/39 | 43/55 | 6/8 |
| 中部偏东南 | F | 0/0 | 2/9 | 3/8 | 13/36 | 8/19 | 15/24 | 5/5 |
| | M | 0/1 | 2/6 | 2/8 | 11/25 | 13/16 | 9/16 | 2/6 |
| 中部偏西南 | F | 0/0 | 3/10 | 7/20 | 19/40 | 10/19 | 17/22 | 4/5 |
| | M | 0/1 | 1/4 | 5/13 | 20/47 | 13/16 | 19/25 | 0/3 |
| 落基山区 | F | 0/0 | 1/5 | 4/14 | 22/41 | 11/14 | 14/16 | 4/5 |
| | M | 0/0 | 3/3 | 4/12 | 16/27 | 16/22 | 20/24 | 2/2 |
| 太平洋地区 | F | 1/4 | 7/29 | 7/19 | 40/79 | 23/33 | 29/37 | 3/4 |
| | M | 0/1 | 4/21 | 8/25 | 21/40 | 33/41 | 22/25 | 3/4 |

数据来源：1982 年、1983 年、1984 年、1985 年的综合社会调查项目（GSS）。

注：政治态度取值为从“1 = 极端自由义”到“7 = 极端保守主义”。每个单元格中包括两个数目：投票选里根的人数/总人数。

数据中没有关于“大城市统计标准区域（Standard Metropolitan Statistical Areas）”的信息，因此，就用居住区域来表示。同一区域中的观察可能行为更为一致，必然导致区域内部的相关，这便引起了离散异质性问题。非恒定的离散可能会进一步影响对分组的参数相同性检验。为了评估离散异质问题的复杂性，我们再次分析表 6.1 中的数据。现在数据被按照地区拆分，如表 7.3 所示。我们只列出了九个区域来演示，实际上区域的数目要比这多得多。

注意，表 7.3 中有四个观察中总投票人数为零。在以下的小节中这些个案将从分析中忽略掉，因为我们要使用对数单位联系来为数据建模。对于其他联系函数来说，频次为零并不是一个问题。例如，对于恒等联系来说，取值为零的个案并不构成问题。

## 7.3.2　存在异质离散时的分组比较

为了评估离散异质性对于参数相同性检验的影响，我们将考察三个模型。第一个模型是传统的 GLM。它是一个“原生模型（naive model）”，因为它假定 $\phi$ 是恒定的。为了简化对各种离散异质性方法的比较，我们使用包括与分组变量交互作用的单模型检验。第二个模型假定 $\phi$ 随组别而变动，用权重 $\hat{\phi}_g$ 对标准误进行调整，权重 $\hat{\phi}_g$ 可以是 GLM 拟合中的量纲权重，也可以是应用于 y 向量的数据权重。第三个模型进一步放宽了假定，用一个对数联系将 $\phi$ 设定为某些 z 变量的函数。表 7.4 介绍了这三个模型的结果，以及基于这三个模型而进行的各组参数相同性的检验结果（LRT 和 QLRT）。

**表 7.4　对选举数据的对数单位模型的 GLM 和 DGLM 估计值**

| | GLM | | | DGLM | | | |
|---|---|---|---|---|---|---|---|
| x×z 变量 | $\hat{\beta}$ | $se_1$ | $se_2$ | $\hat{\beta}$ | se | $\hat{\lambda}$ | se |
| 截距 | -2.347 | 0.217 | 0.282 | -2.494 | 0.253 | -0.208 | 0.320 |
| 是否女性 | 0.384 | 0.290 | 0.329 | 0.502 | 0.295 | -0.752 | 0.256 |
| 政治态度 | 0.595 | 0.049 | 0.063 | 0.630 | 0.060 | 0.143 | 0.065 |
| 是否女性×政治态度 | -0.132 | 0.066 | 0.074 | -0.160 | 0.070 | — | — |
| $\beta_m=\beta_f$ 的（Q）LRT | — | 9.188 | 7.414 | | | 8.281 | |
| （Q）LRT 的自由度 | — | 2 | 2 | | | 2 | |

注：$se_1$ 表示从 GLM 中得到的未修正标准误，$se_2$ 表示从用分组的 $\phi_g$ 估计量修正过的 GLM 中得到的标准误。（Q）LRT 用来检验不同性别间截距项和政治态度的参数的差异。

由于只有标准误才用 $\hat{\phi}_g$ 进行调整，故前两个模型有相同的参数估计。使用式（7.4），可以算出男性的 $\hat{\phi}_g$ 为 1.691 3，女性的 $\hat{\phi}_g$ 为 0.763 9；包括两性的

合并模型的 $\hat{\phi}_g$ 为 1.239 8。这表明不论在合并模型中存在过度离散与否，两组间都存在着相当大的离散异质性。使用式(7.5)比较男性和女性的 $\chi_g^2$，得到 F 比为 $2.214_{[60,58]}$，它在 0.01 水平下显著。使用式(7.6)中的一般 F 比来检测离散异质性，得到结果为 $11.038_{[4,118]}$，它在 0.01 水平下也是显著的。这样，每一种 F 检验都表明违反了两组间离散同质性的假定。两个 $\phi_g$ 估计作为量纲权重被用来重新估计两个模型。这种调整并未改变参数估计，只是改变了标准误。调整后的标准误被标注为 $se_2$，而相邻的未加调整的名义标准误被标注为 $se_1$。但是，这并未影响到实质结论，因为所有的参数估计(除了性别变量的参数以外)与其标准误相比都很大，无论是名义上的还是调整后的。但是，就总体而言，调整后的标准误都比非调整的标准误大。在调整之后，离散参数估计都是 1，这是假定中已经设定好了的。

由于存在离散异质性，第 6 章中介绍的 LRT 就不能充分地评估参数相同性假定了。因此，我们转而使用基于准似然的检验方法。首先，进行式(7.7)中的准似然比检验，计算全模型(full model)的残余偏差测量与去掉涉及性别变量的两项之后的模型的残余偏差测量之差值。得到的 QLRT 统计量为 7.414，自由度为 2。这一结果表明，不论两性间是否有相同的离散，两组参数估计之差异的显著度在调整之后要比调整之前更低，虽然在 0.05 水平下仍然显著。

表 7.4 中 DGLM 下的四列分别代表 β 和 λ 估计量，以及它们各自在均值子模型和离散子模型中的标准误。可以看出，β′的标准误都比离散同质性假定下的“原生模型”中的标准误大，但比用 $\hat{\phi}_g$ 加权的模型中的标准误小。在 DGLM 中，φ 被视为两个 z 变量的函数：性别和政治态度。λ 与标准误之比表明这两个变量都造成了异质的离散。由于向量 β 是用更新后的 λ 值进行迭代估计的，因此，向量中的最终估计值不同于未加权或加权后的 GLM 中的估计值。

用来检验不同性别间参数相同性的假设 $\beta_m = \beta_f$ 的 QLRT 统计量是通过拟合两个 DGLM 得到的。第一个 DGLM 的均值子模型中只有一个解释变量“政治态度”，而在离散子模型中有两个解释变量“性别”和“政治态度”。第二个 DGLM 的离散子模型设定与第一个 DGLM 相同，而在均值子模型中包括了“性别”变量的主效应和它与“政治态度”的交互作用项，如表 7.4 所示。这两个模型的残差测量的差值就给出了式(7.7)中的检验。正如 DGLM 的均值子模型中的标准误一样，DGLM 的 QLRT 统计量也落在了“原生模型”和用 $\hat{\phi}_g$ 调整过的 GLM 模型之间。

在研究参数比较时，我们发现离散异质性不仅影响标准误和参数相同性的检验统计量，而且还可能改变均值子模型中的 β 估计。因此，即使建模的目标不是为了进行分组比较，我们仍然应当警惕离散异质性问题。以下五个步骤的程序有助于组织这项工作：

1. 对原始的未分组数据计算 $\phi$。在二项数据和泊松数据中,如果结果偏离了名义值,那就需要进一步来研究数据。但是,$\phi$ 近似于名义值并不一定意味着离散是同质的,因为各组之间的变异可能在总的 $\phi$ 估计中彼此抵消了。

2. 确定哪一个 z 变量影响着 $\phi$ 的变异。如果你对此不确定,可以在离散子模型中先加入你怀疑有影响的变量试验一下,最后再简化离散子模型。

3. 如果研究者想检验 $\beta$ 是否在各组间相同,那就拟合两个 DGLM:一个在均值子模型中包括分组变量及其与 x 变量的交互作用,另一个的均值子模型则不包括;但两个 DGLM 的离散子模型要保持一致。

4. 如果研究者对于组间 $\lambda$ 的相同性假设也感兴趣,也需要拟合两个 DGLM:一个在离散子模型中包括分组变量及其与 z 变量的交互作用,另一个的离散子模型不包括;但两个 DGLM 的均值子模型要保持一致。

5. 如果研究者对于 $\beta$ 和 $\lambda$ 的相同性假设都感兴趣,那就拟合两个 DGLM:一个在均值子模型中包括分组变量及其与 x 变量的交互作用,在离散子模型中包括分组变量及其与 z 变量的交互作用;另一个在均值子模型和离散子模型中都不包括上述几项。不过要记住,在这些的联合检验中统计量显著的原因可能是两种子模型中的其中一种存在分组差异。也就是说,要确定各组之间差异的来源,第 5 步必须与第 3 步和第 4 步一起执行。

上面的例子中只演示了前三步来检验 $\beta$ 的相同性,但是上述方法可以很容易地扩展到对 $\lambda$ 在各组之间的相同性检验中,特别是在 S-plus 中可以很方便地执行 DGLM 功能。

## 7.4　贝叶斯广义线性模型

近年来,贝叶斯统计已经从一种边缘性方法发展成为现代应用统计学和理论统计学的支柱之一。GLM 模型也受到了这种发展趋势的影响。贝叶斯法是一种具有普适性的统计方法,因此有许多方式对 GLM 进行改进,这一点在《广义线性模型:贝叶斯视角》(Generalized Linear Models: A Bayesian Perspective)一书中得到了反映(Dey, Ghosh, and Mallick, 2000)。在本节中,我们主要关注其中一种方法,它使得贝叶斯的思维方式有助于 GLM 中的分组比较。但是,我们首先要回顾一下贝叶斯统计学的一些基本知识。

### 7.4.1　贝叶斯推断

贝叶斯推断允许研究者把主观性或者信念(先验概率)引入概率计算中,然后利用数据信息(通过似然函数)来进行推断(后验概率)。令 $P(H|B)$ 代表某事件发生的主观概率或信念概率,即在先验信息或信念前提下假设 H 的概率,这被称为**先验概率**。令 $P(D|H,B)$ 代表**似然**,它是在假设 H 和先验信息或信念 B 为真的前提下数据的概率,令 $P(H|D,B)$ 代表考虑了数据和先验

知识之后假设 H 的概率;这被称为**后验概率**。贝叶斯准则或定理为:

$$P(H \mid D,B) = \frac{P(H \mid B)P(D \mid H,B)}{P(D \mid B)}$$

其中,$P(D \mid B)$表示在给定先验信息前提下数据的概率,它被视为是正态化常数(normalizing constant)或量纲常数(scaling constant),因为它与将要考虑的假设或模型无关。因此,一个假设的后验概率与先验概率和似然的乘积成比例。

上式中的所有概率都是条件概率,我们也可以有非条件概率。令 $H_p$ 代表有关事件或数据的所有完备且互斥的假设,然后再从概率中略去 B,因为这并不影响理解,那么则有下式:

$$P(D) = \sum_p P(D \mid H_p)P(H_p)$$

这被称为广义相加法则。另外一个法则是广义相乘法则:

$$P(H_1H_2\cdots H_P) = P(H_1)P(H_2 \mid H_1)P(H_3 \mid H_1H_2)\cdots P(H_P \mid H_1H_2\cdots H_{P-1})$$

从广义相加法则中推导出的重要结果就是贝叶斯定理:

$$P(H_p \mid D)P(D) = P(D,H_p) = P(H_p)P(D \mid H_p)$$

只要 $P(D) \neq 0$,则有下式:

$$P(H_p \mid D) \propto P(H_p)P(D \mid H_p) \tag{7.11}$$

这是贝叶斯定理的众多表达式之一,它比在分母中包括 $P(D)$作为量纲常数的表达式更为简化。

如果 $H_p$ 代表所有完备且互斥的假设 1 到 P,从贝叶斯定理可导出下式:

$$P(H_p \mid D) = \frac{P(H_p)P(D \mid H_p)}{\sum_{p=1}^{P} P(H_p)P(D \mid H_p)} \tag{7.12}$$

#### 7.4.1.1 示例

为了演示贝叶斯定理的简单应用,我们来看一项关于美国人对于堕胎药的态度的研究。2000 年 9 月,美国食品和药品署(以下简称 FDA)批准销售一种法国产的堕胎药 RU-486。民主党人对于合法堕胎的典型立场是"可以",民主党的总统候选人戈尔就是这样表态的;共和党人对于这一问题的典型立场是"不可以,除非有极端情况",共和党的总统候选人布什的表态与此一致。候选人的态度在 2000 年 9 月大选的争论中得到了澄清。假定民主党(简称 DM)以 53% 的民众支持率赢得了选举,而共和党(简称 RP)获得了 47% 的民众支持率(为了简单起见假定其他候选人的支持率可以忽略不计)。再假定在大选之后,一项民意调查显示有 54% 的共和党投票者和 86% 的民主党投票者支持克林顿政府将 RU-486 在美国合法化的决定。最后,假定你当时遇到了某人支持 FDA 批准销售 RU-486。你现在想猜一猜他支持的是哪个党派。给定上述信息,依据贝叶斯定理,则:

$$P(DM \mid Y) = \frac{P(Y \mid DM)P(DM)}{P(Y \mid DM)P(DM) + P(Y \mid RP)P(RP)}$$

$$= \frac{(0.86)(0.53)}{(0.86)(0.53) + (0.54)(0.47)} = 0.64$$

因此,此人在大选中支持了民主党的概率是0.41。

#### 7.4.1.2　贝叶斯模型比较

假定我们有两个相互竞争的假设:$H_p$, $p = 0,1$。和以前一样,$H_0$ 代表两组观察中 $\beta_1 = \beta_2$。依据贝叶斯定理(7.12),则:

$$P(H_p \mid D) = \frac{P(D \mid H_p)P(H_p)}{P(D \mid H_0)P(H_0) + P(D \mid H_1)P(H_1)} \quad \text{其中 } p = 0,1 \qquad (7.13)$$

其中,数据 D 被认为是在两种假设中某一种假设成立的前提下出现的,这两种竞争假设 $H_0$ 和 $H_1$ 有着各自的概率密度 $P(D \mid H_0)$ 和 $P(D \mid H_1)$,而 $P(H_p \mid D)$ 是数据给定情况下 $H_p$ 的后验概率。$H_1$ 的后验概率与 $H_0$ 的后验概率之比就是后验几率(posterior odds),它测量了数据支持 $H_1$ 胜过支持 $H_0$ 的程度:

$$\frac{P(H_1 \mid D)}{P(H_0 \mid D)} = \frac{P(D \mid H_1)}{P(D \mid H_0)} \frac{P(H_1)}{P(H_0)} \qquad (7.14)$$

其中,右边的第一个因子被称为贝叶斯因子,第二个因子被称为先验几率。不同于第 7.1.4.1 节中关于堕胎药的例子,我们常常没有先验知识来判断某一个假设比另一个假设更有可能性,因此只能给每个先验概率相等的赋值。就这两个相互竞争的假设来说,先验几率就是0.5/0.5 =1。

为了得到式(7.14)的后验几率来比较两个假设,我们只需了解似然比,拉弗特里(Raftery,1996)将式(7.13)和式(7.14)中出现的似然定义为:

$$P(D \mid H_p) = \int P(D \mid \theta_p, H_p)P(\theta_p \mid H_p)d\theta_p$$

其中,$\theta_p$ 代表假设 p 中的一些参数。拉弗特里(Raftery,1996)提议用积分中的拉普拉斯法(Laplace method)来近似这里定义的似然。存在三种方法来近似贝叶斯因子(Raftery,1996),这里介绍其中最精确的方法和最简单的方法。

最精确的近似方法依赖于拉普拉斯法,它将贝叶斯因子的自然对数的两倍表达为:

$$2 \ln B_{10} \approx \chi^2_{10} + (E_1 - E_0) \qquad (7.15)$$

其中,$B_{10}$表示比较 $H_1$ 和 $H_0$ 的贝叶斯因子,代表 $H_0$ 的模型嵌套于代表 $H_1$ 的模型之中,$\chi^2_{10}$是分别代表 $H_1$ 和 $H_0$ 的模型的似然比 $\chi^2$ 之差值。$E_p$($p = 1,0$)是对模型 p(即假设 p 下的模型)在 $\hat{\theta}_p$ 与 $E(\theta_p \mid H_p)$ 的接近程度上的表现的二次测度(quadratic measures)。如果想了解更多的计算细节,请参阅拉弗特里的著作(Raftery,1996:254)。

最简单的近似方法是贝叶斯信息准则(以下简称 BIC),它是对 LRT 的直

接而简单的调整：

$$2 \ln B_{1\text{-}0} \approx \chi^2_{1\text{-}0} - df_{1\text{-}0} \ln N \tag{7.16}$$

其中，$df_{1\text{-}0}$是代表假设1的模型的自由度和代表假设0的模型的自由度之差值。$H_0$既可以是基准模型（如饱和模型），也可以是对自变量参数进行约束的零模型。如果基准模型是饱和模型，那么，用到的LRT就是一个**偏差**测量。如果基准模型是一个零模型（即没有任何自变量的模型），BIC的公式就变为：

$$2 \ln B_{1\text{-}0} \approx -\chi^2_{1\text{-}0} + df_{1\text{-}0} \ln N$$

这保持了BIC在模型比较方面的一致性。式（7.16）中给出的统计量也被称为“施瓦茨准则”（Schwarz's criterion）。BIC也可用于彼此并不嵌套的模型比较中。

但是，$H_0$和$H_1$只是所有可能模型中的两个模型，检验各组之间的参数相同性比这种比较要涉及更多的内容。本章中我们将考察在多组之间进行贝叶斯模型比较的方法，这种方法是由我本人给出的（Liao，2002）。例如，对于每个$\beta_p$中包括两个参数$\beta_{p1}$和$\beta_{p2}$的零假设$\beta_1 = \beta_2$，标准的备择假设$\beta_1 \neq \beta_2$其实包含了三个子假设：$\beta_{11} \neq \beta_{21}$；$\beta_{12} \neq \beta_{22}$；$\beta_{11} \neq \beta_{21}$且$\beta_{12} \neq \beta_{22}$。这意味着在模型空间中，总共有四个模型，其中一个模型代表$H_0$，另外三个模型代表$H_1$。

这个简单的例子表明：我们在进行分组比较时需要允许模型有不确定性，需要考虑相关模型空间中的所有可能的合理模型，并且相应地定义和计算它们的后验概率。假定针对零假设$H_0$要考察的模型共有P个：$H_1, \cdots, H_P$。每个假设都是针对基准假设$H_0$而检验，这就产生了贝叶斯因子$B_{1\text{-}0}, B_{2\text{-}0}, \cdots, B_{P\text{-}0}$（Kass and Raftery，1995）。如果考虑到所有相关模型，那么，$H_p$的后验概率为：

$$P(H_p \mid D) = \frac{\alpha_p B_{p\text{-}0}}{\sum_{p=1}^{P} \alpha_p B_{p\text{-}0}} \tag{7.17}$$

其中，$\alpha = P(H_p)/P(H_0)$是支持$H_p$胜于支持$H_0$的先验几率。一种很自然的选择是把所有的先验几率都设定为1。我们可把式（7.17）重新表示为$2 \ln B_{p\text{-}0}$的函数（Raftery，1995）：

$$P(H_p \mid D) \approx \frac{e^{-0.5 \ln B_{p\text{-}0}}}{\sum_{p=1}^{P} e^{-0.5 \ln B_{p\text{-}0}}} \tag{7.18}$$

为了检验两组间参数相同性的假设，式（7.18）中的加总必须包括模型空间中相关的所有模型。如果检验的是多组之间的参数相同性假设，我们应当包括模型空间中所有涉及相关参数的层次（即嵌套的）模型。

借助于式（7.18），我们可以相当灵活地构建假设来检验各组之间的参数相同性。也就是说，不论是零假设还是备择假设还是它们两者同时都可以是用多个模型表示的组合假设。以检验两组中两个参数的相同性为例，我们的备择假设可以是两组间至少有一对非截距参数不相同。也就是说，我们的备择假设$H_1$包括两个模型：

- $\beta_{10}=\beta_{20}$但是 $\beta_{11}\neq\beta_{21}$；
- $\beta_{10}\neq\beta_{20}$并且 $\beta_{11}\neq\beta_{21}$。

另一方面，我们的零假设也由两个模型构成：

- $\beta_{10}\neq\beta_{20}$但是 $\beta_{11}=\beta_{21}$；
- $\beta_{11}=\beta_{21}$并且 $\beta_{12}=\beta_{22}$。

因此，在这四个模型中，两个模型构成了零假设，两个模型构成了备择假设。

总而言之，由于贝叶斯方法具有的灵活性，我们用以检验各组之间参数相同性的假设 $H_1$ 可以分为三种类型。在此我们用 H 表示一个假设，用 $M_K$ 表示代表此假设的模型。检验相同性的一种方法是使用分组变量与解释变量的交互项。我们采用这种方法来进行广义线性模型在贝叶斯情境下的检验：

1. 假设类型 A($H_A$)：即认为所有组的所有模型参数总体而言有所不同的传统假设。如果要比较的只有两个模型，这就是标准的检验，它可以用 LRT，BIC 或者其他等效的方法进行。如果我们使用虚拟变量交互项的方法，零假设就是模型中只有解释变量的主效应（即没有分组变量）。

2. 假设类型 B($H_B$)：在所有可能的模型集合中，有某一模型子集被认为是零假设，而其余的模型构成了备择假设，零假设中的模型嵌套于备择假设中的模型。下一节中的示例要比较两组中的两个参数，它就是这样一种假设类型。

3. 假设类型 C($H_C$)：这是类型 B 的进一步扩展。在此类型中，在检验组间的参数相同性时，代表零假设的模型($H_0$)并不一定要嵌套于代表备择假设的模型($H_1$)。

要检验诸如 $H_A$ 这样的传统假设，我们要考虑两个模型：$H_0$ 由只包括解释变量主效应的模型代表；$H_1$ 由包括了分组变量与每个解释变量的交互作用的模型代表。如果要检验的是 $H_B$ 这样的假设，那么 $H_1$ 就由所有包括了分组变量与解释变量的任何一项交互作用的模型代表；$H_0$ 由模型空间中所有不包括此种交互作用的模型代表。$H_C$ 则由建模者根据具体情形定义，它可以是不嵌套的。

为了评估这些类型的假设，我们有如下公式来构建后验几率：

$$\frac{P(H_\Omega \mid D)}{P(H_{\bar{\Omega}} \mid D)}=\frac{\sum_{M_k\in\Omega}\alpha_k(M_\Omega)}{\sum_{M_k\notin\Omega}\alpha_k(M_{\bar{\Omega}})}\frac{\sum_{M_k\in\Omega}P(M_k \mid D)}{\sum_{M_k\notin\Omega}P(M_k \mid D)} \tag{7.19}$$

其中，$\Omega$ 是代表备择假设 $H_\Omega$ 的所有模型的集合，$H_\Omega$ 可以是 $H_A$，$H_B$，$H_C$ 三种类型中任何一种。$\alpha_k(M_\Omega)$是模型 k 的先验概率，$\alpha_k(M_\Omega)/\alpha_k(M_{\bar{\Omega}})$是模型 k 支持 $H_\Omega$ 胜过支持 $H_{\bar{\Omega}}$ 的先验几率。右端的第二项因子是加总的（后验）概率比。式(7.19)可理解为用先验几率来调整后验概率比。如果所有的先验概率都相同，则先验几率事实上等于一减去 $\Omega$ 中的模型数目与模型总数目之比。例如，对于有两个自变量和三个组的模型来说，模型空间中共有 47 个模型，其中

31 个代表 $H_B$ 类型的假设 $H_\Omega$，其余 16 个代表 $H_{\bar{\Omega}}$。有着 31 个模型可能会导致“过度代表(overrepresentation)”，因为 Ω 模型有 31 个后验概率加总，而非 Ω 模型中只有 16 个后验概率加总。因此，后验概率和后验几率都要依据先验概率加权，在本例中也就是要依据它们在模型空间中的代表性来加权。类似地，为了检验 $H_C$，我们必须在分子中对 $C_1$ 的后验概率加总，在分母中对 $C_0$ 的后验概率加总，然后再用它们的先验几率进行相应调整。

#### 7.4.1.3 模型空间

为了检验组间的参数相同性，我们的模型空间中包括了所有满足层次原则(hierarchical principle)的模型(即只有在模型中包括了某一高阶交互项的主效应及其低阶交互项时，模型中才会包括这一高阶交互项)。因此，模型空间中包括了所有的模型，从只有截距项的模型，到包括了分组变量和解释变量的所有可能组合的模型，到包括了分组变量和解释变量间交互作用的所有可能组合的模型，只要它们满足层次原则。

依据这种逻辑，模型空间中满足层次原则的模型可分为两组：有交互项的模型和没有交互项的模型。模型的总数(以下简称 NM)为：

$$\mathbf{NM} = \sum_{r=0}^{G+J}\binom{G+J}{r} + \sum_{s=1}^{G}\sum_{t=1}^{J}\left[\binom{G}{s}\binom{J}{t}\sum_{u=1}^{s\times t}\binom{s\times t}{u}\right] + \sum_{r=1}^{G\times J}\binom{G\times J}{r} \qquad \text{其中 } s\times t < G\times J \tag{7.20}$$

其中，$\binom{n}{r}$是从 n 个对象中选取 r 个对象所有组合的个数(即 $n!/[(n-r)!r!]$)。G 是“唯一组对(unique grou Pcontrasts)”的总数，J 是解释变量的总数。例如，如果进行两对组比较，有两个自变量，那么，模型空间中就会存在 47 个模型。

在这里，关键的一个数量是模型空间中“唯一组对”的总数。如果是对定类变量的多个类别进行成对组比较，那么，G 组则存在 G－1 个“唯一组对”。如果模型中包含了所有的“唯一组对”，其形式都是用虚拟变量代表一个类别变量，那么参照类别的选择并不会影响到对类别变量的效应的整体统计检验(如 F 检验或 $\chi^2$ 检验)。与此类似，模型中只要包含了 G－1 个“唯一组对”，那么，参照组的选择也不会影响对各个组之间的参数差异的整体检验。但是，如同对类别变量的处理一样，参照类别或参照组的选择决定了参数代表了哪个类别或哪个组，因此，也影响到对参数的解释。如果研究者对检验组间差异和解释估计值感兴趣，那么为了便于解释，他应当选择解释变量的效应为零(或者接近于零)的那组作为参照组，这样其他组就可解释为对于零效应的相对准则的偏离。

#### 7.4.1.4　软件

阿德里安· 拉弗特里(Adrian Raftery)和他的同事开发了用单步牛顿近似法来估计贝叶斯 GLM 的软件,它是 S 语言函数(glib),可以从卡内基-梅隆大学互联网上的 Statlib 中得到。我本人在 glib 框架基础上进行扩展,使其可以进行多组比较,它也是 S 语言函数,名为 multigroup. glib,可以从如下网址得到:http://www. staff. uiuc. edu/ ~ f-liao/Statcompare/multigroup. glib。

这一程序的输出格式与 glib 的输出格式基本一致,只是增加了两个变量:G(组对的数目);K(自变量的数目,这是为了与 glib 的标注相一致,在 glib 中 j 代表其他意义)。如果假设为 $H_A$ 和 $H_B$ 类型,输出格式中还增加了后验概率和后验几率的估计。对 $H_C$ 类型的检验要取决于具体情况和用户自己的定义。输出中包括了模型空间中所有模型的后验概率,用户可以自己来构建 $H_C$ 检验的后验概率和后验几率。

或者,数据分析者也可以使用标准的统计软件计算基于 LRT 的 BIC。只要变量总数不太多,模型空间中的模型不太多,这种方法就是可行的。

### 7.4.2　示例

本节中将以我的一项研究为例(Liao,2002)来说明如何进行两组间多个自变量的参数差异检验。这里要考虑的实质性问题是自我评价与学业成就之间的关系。学业成就和性别对于自我评价有无影响呢?如果有影响,那么这种影响在黑人大学生和白人大学生之间有无差别呢?这些问题可以用迪莫和帕克的数据来解答(Demo and Parker,1987)。分组变量是种族。这一决策虽然略显武断,但是在理论上是站得住脚的,在教学上也很有用。表 7.5 中介绍了这项数据。由于结果是二项变量,所以我们使用了对数单位联系。因此,我们采用两个含有对数单位联系的 GLM 模型来拟合数据,一个代表零模型,另一个代表全模型(以种族来分层)。这样,包括两个种族的不分组模型为:

$$\ln\left[\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right] = \beta_0 + \beta_1 \mathrm{Sex}_i + \beta_2 \mathrm{GPA}_i$$

**表 7.5　大学生中的自我评价与学业成绩(频次表)**

| 性别 | 成绩 | 黑人 | | 白人 | |
|---|---|---|---|---|---|
| | | 高自我评价 | 低自我评价 | 高自我评价 | 低自我评价 |
| 男 | 高 | 15 | 9 | 17 | 10 |
| | 低 | 26 | 17 | 22 | 26 |
| 女 | 高 | 13 | 22 | 22 | 32 |
| | 低 | 24 | 23 | 3 | 17 |

资料来源:Demo and Parker(1987)。

此模型也可分别应用于黑人和白人的观察中，这样的模型就是非限定模型。或者，非限定模型也可采用包括虚拟变量（代表分组）的交互项的形式：

$$\ln\left[\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right] = \beta_0 + \beta_1 Race_i + \beta_2 Sex_i + \beta_3 GPA_i + \beta_4 Race_i \times Sex_i + \beta_5 Race_i \times GPA$$

这两个模型构成了检验两组之间参数相同性的 A 类型的传统假设，第一个模型代表零假设，包括了与"种族"变量的交互项的第二个模型代表备择假设。如表 7.6 所示为这两个模型的 ML 估计。

**表 7.6　利用自我评价数据进行的最大似然估计**

| x-变量 | 不分组 | 非限定模型 | |
|---|---|---|---|
| | | 黑人 | 白人 |
| 截距 | 0.195 | 0.560 | -0.250 |
| | (0.190) | (0.281) | (0.271) |
| 性别（女性=1） | -0.708 | -0.635 | -1.141 |
| | (0.242) | (0.336) | (0.389) |
| 成绩（高=1） | 0.169 | -0.287 | 0.941 |
| | (0.242) | (0.341) | (0.391) |
| LRT：模型 $\chi^2$ | 8.730 | 18.432 | |
| 自由度 | 3 | 6 | |
| BIC | -8.362 | -15.751 | |
| 2 log $B_{10}$（单步牛顿） | | | |
| $\phi=1$ | -2.786 | -7.927 | |
| $\phi=1.65$ | -4.766 | -12.821 | |
| $\phi=5$ | -9.189 | -23.848 | |

注：括号内的值是参数估计的标准误。

$\phi$ 的三个值给出了先验分布的标准差的合理取值范围。

在"不分组"列中的模型是限定模型，它假定性别和成绩的影响在不同种族群体中不存在任何差异；后两列中的模型是非限定模型，它假定 x 的影响因不同种族而不同。两个模型中一个假定黑人学生和白人学生的参数相同，一个假定参数不相同，但它们的 LRT 都给出了一个在任何传统的 $\alpha$ 水平下都显著的 $\chi^2$ 统计量。两个 LRT 的差值给出了另一个 LRT，它可以检验两个模型是否有差异。两个模型估计的参数数目相差 3 个，因此 LRT 自由度为 3，在此自由度下，LRT 是显著的。两个模型的 BIC 差值为 7.4，据此可以比较有把握地说，两个组之间差异的非限定模型更优一些。

在拟合贝叶斯 GLM 时，参数 $\phi$ 代表了先验几率的参照集的合理取值范围。具体来讲，在本例中，依据先验分布的标准差，$\phi$ 的合理取值范围为 1～5，而中间点应为$\sqrt{e}\approx 1.65$（Raftery，1996）。在 $\phi$ 的所有水平上，使用了更为精确的单步牛顿法的每对 $2\ \ln B_{10}$ 估计值都表明，非限定模型对数据的拟合更好。就绝对值而言，BIC 值要比 $\phi=1.65$ 时的单步牛顿估计值高一些。两个模型的 BIC 差值近似于两个模型在 $\phi=1.65$ 时用单步牛顿法估计的 $2\ \ln B_{10}$ 差值。但是，这两个模型只代表了模型空间中 13 个可能模型中的两个模型。表 7.7 中列出了自我评价数据的模型空间中的所有模型。

**表 7.7　自我评价数据的各个模型的设计矩阵**

| 模型 | 种族 | 性别 | 成绩 | 种族×性别 | 种族×成绩 |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 1 | 0 | 0 |
| 5 | 1 | 1 | 0 | 0 | 0 |
| 6 | 1 | 0 | 1 | 0 | 0 |
| 7 | 0 | 1 | 1 | 0 | 0 |
| 8 | 1 | 1 | 1 | 0 | 0 |
| 9 | 1 | 1 | 0 | 1 | 0 |
| 10 | 1 | 0 | 1 | 0 | 1 |
| 11 | 1 | 1 | 1 | 1 | 0 |
| 12 | 1 | 1 | 1 | 0 | 1 |
| 13 | 1 | 1 | 1 | 1 | 1 |

模型空间（$P=13$）中的 13 个模型都服从层次原则，都可用来检验自我评价数据中两个种族群体的参数是否相同。从表 7.7 中我们看出，表 7.6 中介绍的 A 类型的传统假设只是 13 个可能模型中的两个（模型 7 和模型 13）。如果模型空间中只存在这两个模型，我们可以采用式（7.18）计算出它们各自的后验概率。结果是，非限定模型的后验概率为 0.982 5，限定模型的后验概率为 0.017 5，后验几率为 56。结果很有力地拒斥了参数相同假设。卡斯和拉弗特里（Kass and Raftery，1995）对于后验概率和后验几率的解释提出了一些指导原则，表 7.8 中汇总了这些原则。

表 7.8 进行解释和决策的一些准则

| $B_{1-0}$ | $2\ln B_{1-0}$ | 对 $H_1$ 的支持程度 |
|---|---|---|
| <1 | <0 | 不支持 |
| 1~3 | 0~2.2 | 仅仅值得一提 |
| 3~12 | 2.2~5 | 支持 |
| 12~150 | 5~10 | 强烈支持 |
| >150 | >10 | 绝对支持 |

资料来源:Kass and Raftery(1995)。

除了传统的 $H_A$ 假设外,让我们再来关注一下 $H_B$ 和 $H_C$ 假设。第一个类型 $H_B$ 是复合假设,它假设"性别"的影响或者"成绩"的影响或者这两个变量的影响在两个种族群体中存在差异,而相应的零假设认为不存在差异。对应于第二个类型 $H_C$ 的零假设也是复合假设,不论"成绩"在模型中是作为主效应还是与分组变量的交互作用,它认为"成绩"变量没有影响。

为了检验 $H_B$,我们把表 7.7 中后五个模型的后验概率进行加总来代表 $H_B$ 的后验概率,而把其他模型的后验概率也进行加总来代表零假设的后验概率,然后比较这两者。两个后验概率及由此得到的后验几率都要用它们的先验几率按照式(7.19)进行调整。也就是说,五个模型对八个模型的几率和后验概率都要进行重新量纲化(rescaled),使其加总为 1。换而言之,代表五个模型的加总后验概率要乘以 1-5/13 进行加权,而代表八个模型的加总后验概率要乘以 1-8/13 进行加权。对 $H_C$ 的检验中后验概率的计算也与此类似,要对涉及"成绩"变量的模型(即模型 4,6-8,10-13)的后验概率进行加总,也要对不涉及它的模型的后验概率进行加总(即模型 1-3,5,9),然后对其进行相应加权。

让我们用一个简单的例子来说明加权调整和的重新量纲化的必要性。假定我们有十个模型,其中三个代表零假设,其余七个代表备择假设。再假定每个模型都具有相同的概率 0.10。如果不进行加权,零假设的加总后验概率为 0.30,备择假设的加总后验概率为 0.70;由于每个假设所拥有的模型数不相同,故它们也不是相同的。如果我们对 0.30 用 7/10 加权,对 0.70 用 3/10 加权,那么我们将得到零假设的概率为 0.21,备择假设的概率也是 0.21,这样两者的后验概率就又相同了:模型数目不同并没有影响到最终的后验概率。然后得到的概率可以进行重新量纲化,使其加总为 1,即 0.21/(0.21+0.21)=0.50,诸如此类。如果出于某种考虑,研究者希望模型数目能够得到体现,那么就可以选择不去对后验概率进行加权。

表 7.9 中介绍了假设 $H_A$，$H_B$，$H_C$ 的后验概率和后验几率。表 7.9 中的后验概率和后验几率都是针对备择假设而言的，即假设两个群体间存在差异。从后验概率和后验几率来判断，不论它采用了传统的 $H_A$ 形式还是复合的 $H_B$ 形式，两个种族群体间存在差异的假设都得到了支持。但是，$H_B$ 考虑了模型空间中所有相关的模型，因而它虽然也拒斥了两组之间差异不存在的假设，但给出的证据比 $H_A$ 显得更为谨慎(后验几率范围为 5 ~ 109)。$H_C$ 认为"成绩"要么作为解释变量要么与"种族"变量构成交互作用产生了影响；它从数据中得到的支持是微弱的(后验几率的范围为 0.7 ~ 5.5)。

**表 7.9　对表 7.5 中数据的贝叶斯后验信息**

| | 后验信息 | | |
|---|---|---|---|
| | $\phi=1$ | $\phi=1.65$ | $\phi=5$ |
| 后验概率： | | | |
| $H_A$ | 0.928 9 | 0.982 5 | 0.999 3 |
| $H_B$ | 0.833 3 | 0.921 4 | 0.990 9 |
| $H_C$ | 0.420 8 | 0.531 8 | 0.845 8 |
| 后验几率： | | | |
| $H_A$ | 13.073 7 | 56.112 4 | 1 524.490 9 |
| $H_B$ | 4.998 1 | 11.714 8 | 109.0996 |
| $H_C$ | 0.726 4 | 1.136 0 | 5.483 5 |

我们必须认识到，这个例子只是显示了贝叶斯框架在进行分组比较时通过考虑整个模型空间而有助于捕捉到模型的不确定性。除了为模型空间中的模型计算了成比例的权重，在本例中我们并没有显示出配置先验概率的灵活性，但我们是完全可以通过用户自定义先验权重而把先验知识考虑进来的，这在本节中讨论过的贝叶斯 GLM 软件中是非常容易做到的。

# 7.A　用于 n∶m 设计中的数据

如表7.10 所示为在 7.2.3 节中讨论 n∶m 设计时所使用的数据。

表 7.10　用 n：m 设计的“低出生体重”(LBW)数据

| 年龄 | LBW | WLM | SDP | HT | UI |
|---|---|---|---|---|---|
| 16 | 1 | 130 | 0 | 0 | 0 |
| | 0 | 110 | 0 | 0 | 0 |
| | 0 | 112 | 0 | 0 | 0 |
| | 0 | 135 | 1 | 0 | 0 |
| | 0 | 135 | 1 | 0 | 0 |
| | 0 | 170 | 0 | 0 | 0 |
| | 0 | 95 | 0 | 0 | 0 |
| 17 | 1 | 130 | 1 | 0 | 1 |
| | 1 | 110 | 1 | 0 | 0 |
| | 1 | 120 | 1 | 0 | 0 |
| | 1 | 120 | 0 | 0 | 0 |
| | 1 | 142 | 0 | 1 | 0 |
| | 0 | 103 | 0 | 0 | 0 |
| | 0 | 122 | 1 | 0 | 0 |
| | 0 | 113 | 0 | 0 | 0 |
| | 0 | 113 | 0 | 0 | 0 |
| | 0 | 119 | 0 | 0 | 0 |
| | 0 | 119 | 0 | 0 | 0 |
| | 0 | 120 | 1 | 0 | 0 |
| 18 | 1 | 148 | 0 | 0 | 0 |
| | 1 | 110 | 1 | 0 | 0 |
| | 0 | 107 | 1 | 0 | 1 |
| | 0 | 100 | 1 | 0 | 0 |
| | 0 | 100 | 1 | 0 | 0 |
| | 0 | 90 | 1 | 0 | 1 |
| | 0 | 90 | 1 | 0 | 1 |
| | 0 | 229 | 0 | 0 | 0 |
| | 0 | 120 | 1 | 0 | 0 |
| | 0 | 120 | 0 | 0 | 0 |
| 19 | 1 | 91 | 1 | 0 | 1 |
| | 1 | 102 | 0 | 0 | 0 |
| | 1 | 112 | 1 | 0 | 1 |
| | 0 | 182 | 0 | 0 | 1 |
| | 0 | 95 | 0 | 0 | 0 |
| | 0 | 150 | 0 | 0 | 0 |
| | 0 | 138 | 1 | 0 | 0 |
| | 0 | 189 | 0 | 0 | 0 |
| | 0 | 132 | 0 | 0 | 0 |
| | 0 | 115 | 0 | 0 | 0 |
| | 0 | 105 | 0 | 0 | 0 |
| | 0 | 235 | 1 | 1 | 0 |
| | 0 | 147 | 1 | 0 | 0 |
| | 0 | 147 | 1 | 0 | 0 |
| | 0 | 184 | 1 | 1 | 0 |
| | 0 | 120 | 1 | 0 | 0 |
| 20 | 1 | 150 | 1 | 0 | 0 |
| | 1 | 125 | 0 | 0 | 1 |
| | 0 | 120 | 0 | 0 | 0 |
| | 1 | 80 | 1 | 0 | 1 |
| | 1 | 109 | 0 | 0 | 0 |
| | 1 | 121 | 1 | 0 | 1 |
| | 1 | 122 | 1 | 0 | 0 |
| | 1 | 105 | 0 | 0 | 0 |
| | 0 | 105 | 1 | 0 | 0 |
| | 0 | 120 | 0 | 0 | 1 |
| | 0 | 103 | 0 | 0 | 0 |
| | 0 | 169 | 0 | 0 | 1 |
| | 0 | 141 | 0 | 0 | 1 |
| | 0 | 121 | 1 | 0 | 0 |
| | 0 | 127 | 0 | 0 | 0 |
| | 0 | 120 | 0 | 0 | 0 |
| | 0 | 170 | 1 | 0 | 0 |
| | 0 | 158 | 0 | 0 | 0 |
| 21 | 1 | 165 | 1 | 1 | 0 |
| | 1 | 200 | 0 | 0 | 1 |
| | 1 | 103 | 0 | 0 | 0 |
| | 1 | 100 | 0 | 0 | 0 |
| | 1 | 130 | 1 | 1 | 0 |
| | 0 | 108 | 1 | 0 | 1 |
| | 0 | 124 | 0 | 0 | 0 |
| | 0 | 185 | 1 | 0 | 0 |
| | 0 | 160 | 0 | 0 | 0 |
| | 0 | 110 | 1 | 0 | 1 |
| | 0 | 134 | 0 | 0 | 0 |
| | 0 | 115 | 0 | 0 | 0 |
| 22 | 1 | 130 | 1 | 0 | 1 |

续表

| 年龄 | LBW | WLM | SDP | HT | UI | 年龄 | LBW | WLM | SDP | HT | UI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 1 | 130 | 1 | 0 | 0 | 25 | 1 | 105 | 0 | 1 | 0 |
| | 0 | 118 | 0 | 0 | 0 | | 1 | 85 | 0 | 0 | 1 |
| | 0 | 95 | 0 | 1 | 0 | | 1 | 115 | 0 | 0 | 0 |
| | 0 | 85 | 1 | 0 | 0 | | 1 | 92 | 1 | 0 | 0 |
| | 0 | 120 | 0 | 1 | 0 | | 1 | 89 | 0 | 0 | 0 |
| | 0 | 130 | 1 | 0 | 0 | | 1 | 105 | 0 | 0 | 0 |
| | 0 | 158 | 0 | 0 | 0 | | 0 | 118 | 1 | 0 | 0 |
| | 1 | 112 | 1 | 0 | 0 | | 0 | 120 | 0 | 0 | 1 |
| | 0 | 131 | 0 | 0 | 0 | | 0 | 155 | 0 | 0 | 0 |
| | 0 | 125 | 0 | 0 | 0 | | 0 | 125 | 0 | 0 | 0 |
| | 0 | 169 | 0 | 0 | 0 | | 0 | 140 | 0 | 0 | 0 |
| | 0 | 129 | 0 | 0 | 0 | | 0 | 95 | 1 | 0 | 1 |
| 23 | 1 | 97 | 0 | 0 | 1 | | 0 | 241 | 0 | 1 | 0 |
| | 1 | 187 | 1 | 0 | 0 | | 0 | 120 | 0 | 0 | 0 |
| | 1 | 187 | 1 | 0 | 0 | | 0 | 130 | 0 | 0 | 0 |
| | 1 | 120 | 0 | 0 | 0 | 26 | 1 | 117 | 1 | 0 | 0 |
| | 1 | 110 | 1 | 0 | 0 | | 1 | 96 | 0 | 0 | 0 |
| | 1 | 94 | 1 | 0 | 0 | | 1 | 154 | 0 | 1 | 0 |
| | 0 | 130 | 0 | 0 | 0 | | 0 | 190 | 1 | 0 | 0 |
| | 0 | 128 | 0 | 0 | 0 | | 0 | 113 | 1 | 0 | 0 |
| | 0 | 119 | 0 | 0 | 0 | | 0 | 168 | 1 | 0 | 0 |
| | 0 | 115 | 1 | 0 | 0 | | 0 | 133 | 1 | 0 | 0 |
| | 0 | 190 | 0 | 0 | 0 | | 0 | 160 | 0 | 0 | 0 |
| | 0 | 123 | 0 | 0 | 0 | 27 | 1 | 150 | 0 | 0 | 0 |
| | 0 | 130 | 0 | 0 | 0 | | 1 | 130 | 0 | 0 | 1 |
| | 0 | 110 | 0 | 0 | 0 | | 0 | 124 | 1 | 0 | 0 |
| 24 | 1 | 128 | 0 | 0 | 0 | 28 | 1 | 120 | 1 | 0 | 1 |
| | 1 | 132 | 0 | 1 | 0 | | 1 | 95 | 1 | 0 | 0 |
| | 1 | 155 | 1 | 0 | 0 | | 0 | 120 | 1 | 0 | 0 |
| | 1 | 138 | 0 | 0 | 0 | | 0 | 120 | 0 | 0 | 0 |
| | 1 | 105 | 1 | 0 | 0 | | 0 | 167 | 0 | 0 | 0 |
| | 0 | 90 | 1 | 0 | 0 | | 0 | 140 | 0 | 0 | 0 |
| | 0 | 115 | 0 | 0 | 0 | | 0 | 250 | 1 | 0 | 0 |
| | 0 | 110 | 0 | 0 | 0 | | 0 | 130 | 0 | 0 | 0 |
| | 0 | 115 | 0 | 0 | 0 | 29 | 1 | 130 | 0 | 0 | 1 |
| | 0 | 133 | 0 | 0 | 0 | | 0 | 123 | 1 | 0 | 0 |
| | 0 | 110 | 0 | 0 | 0 | | 0 | 150 | 0 | 0 | 0 |
| | 0 | 110 | 0 | 0 | 0 | | 0 | 140 | 1 | 0 | 0 |
| | 0 | 116 | 0 | 0 | 0 | | 0 | 135 | 0 | 0 | 0 |

续表

| 年龄 | LBW | WLM | SDP | HT | UI |
|---|---|---|---|---|---|
| 29 | 0 | 154 | 0 | 0 | 0 |
| | 0 | 130 | 1 | 0 | 0 |
| | 1 | 142 | 1 | 0 | 0 |
| | 0 | 107 | 0 | 0 | 1 |
| | 0 | 95 | 1 | 0 | 0 |
| | 0 | 153 | 0 | 0 | 0 |
| | 0 | 110 | 0 | 0 | 0 |
| | 0 | 137 | 0 | 0 | 0 |
| | 0 | 112 | 0 | 0 | 0 |
| 31 | 1 | 102 | 1 | 0 | 0 |
| | 0 | 100 | 0 | 0 | 1 |
| | 0 | 215 | 1 | 0 | 0 |
| | 0 | 150 | 1 | 0 | 0 |
| | 0 | 120 | 0 | 0 | 0 |
| 32 | 1 | 105 | 1 | 0 | 0 |
| | 0 | 121 | 0 | 0 | 0 |
| | 0 | 132 | 0 | 0 | 0 |
| | 0 | 134 | 1 | 0 | 0 |
| | 0 | 170 | 0 | 0 | 0 |
| | 0 | 186 | 0 | 0 | 0 |

第 8 章

# 结构方程模型中的比较

Comparison in Structural Equation Modeling

## 8.1 导 论

结构方程模型(以下简称 SEM)作为线性回归分析的一种扩展,已经越来越多地应用于社会科学和行为科学中。SEM 之所以如此受欢迎,主要有两个方面的原因:首先最重要的是,SEM 使数据分析者更容易地检验、确证以及构建理论,因此,它比线性回归和因子分析等传统方法更强些。理论性的重要要求是经验模型能够代表理论模型。线性模型只能依赖于可测量变量,而因子分析的目的在于探索概念与可观察变量之间的关系,而不是探索不可直接测量的概念之间的关系。有些实质理念中存在着潜变量,而潜变量又是由可观察变量来代表的,故结构方程模型很清晰地展现了这类理论。其次,专业软件的出现大大地增强了结构方程模型的实用性。现在,结构方程模型的使用再也不只局限于部分专家,人们可以很便捷地得到 LISREL,EQS,Mplus,Amos 等软件。

在结构方程模型中,包括两类变量:不可观察的潜变量和可观察的指标变量。每一个潜变量只能用可观察变量来测量,它代表着一个实质概念,有一个误差项。之所以要在方程中包括误差项,是因为结构方程模型承认我们对于概念的测量是不完善的。结构方程模型通过联立估计可以灵活地将多个方程包括进一个模型中来,从而进一步增加了模型结构的复杂性,同时也增加了理论的复杂性。

潜变量可以是连续的,也可以是离散的。尽管可观察变量可以不是数量型的,但结构方程模型仍将潜变量设定为连续的。我们将在下一章中考察离散或类别潜变量。在下一节中我们将介绍结构方程模型的统计背景。对此问题不熟悉但希望进一步了解的读者,可以参阅博伦的书(Bollen,1989),这本书全面、清晰、详尽地介绍了结构方程模型的许多问题;也可以参阅米勒的书(Mueller, 1996),这本书重点介绍了模型在 LISREL 和 EQS 中的应用。接下来,我们将介绍均值结构模型,并讨论用结构方程模型进行分组比较。最后,

我们用一个实例来考察如何在结构方程模型中进行均值和协方差比较。

## 8.2 统计学背景

应用结构方程模型的研究包括以下五个步骤：

1. 设定；
2. 识别；
3. 估计；
4. 修正；
5. 解释。

也就是说，数据分析者首先必须设定能够代表研究问题的结构方程模型，并确保模型是可识别的（即所有参数都有唯一的解），同时选择一种方法来估计模型，得出估计结果后若有必要则对模型进行修正，如果模型对数据提供了一个令人满意的拟合还应对结果进行解释。第 4 步通常又会导回到第 1 步。因此，通常拟合一个结构方程模型也可被视为一个迭代过程。我们将在下面简要地讨论这五个步骤。

### 8.2.1 符号与设定

在结构方程模型中，存在两类变量：潜变量和可观察变量。潜在的概念只能通过可观察变量来测量。我们需要结构方程模型的原因在于单个模型中存在多个方程。在一个结构方程模型中，所有以潜变量为自变量而以可观察变量为依变量的方程都被统称为测量模型，而所有以某一潜变量为依变量而有至少一个自变量（无论其是潜变量还是可观察变量）的方程都被统称为潜变量模型（或称为概念模型或构念模型）。有时，不是所有的概念都是潜变量——有些是可以用某一单个指标直接测量的。因此，描述概念之间关系的方程，不论其是由潜变量还是可观察变量表示，都被统称为结构模型。

潜变量可根据它们与方程系统的关系分为两组：如果潜变量在系统中任何地方都不是依变量（即它是事先给定的），那么，它就是外生的，用 $\xi$ 表示；如果潜变量至少在系统中一个方程中是依变量，那么，它就是内生的，用 $\eta$ 表示。例如，我们可能想了解社会经济地位（以下简称 SES）如何影响认知能力。SES，$\xi_1$ 是一个由四个可观察变量 $x_1$ 到 $x_4$ 指示的外生潜变量；而认知能力，$\eta_1$ 是一个由两个能力测试得分 $y_1$ 与 $y_2$ 指示的内生潜变量。因此，在潜变量模型中就只有一个方程，即：

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1$$

其中，$\gamma_{11}$ 是回归系数（第一个下标表示依变量的序号，第二个下标表示自变量的序号），而 $\zeta_1$ 是内生变量 $\eta_1$ 的随机误差。我们假定还有一个潜变量——学业成就，它由社会经济地位和认知能力而决定。这样，在潜变量模型中，则有

另一个方程：

$$\eta_2 = \beta_{21}\eta_1 + \gamma_{21}\xi_1 + \zeta_2$$

其中，$\beta_{21}$ 是 $\eta_2$ 对 $\eta_1$ 进行回归的回归系数。在两个潜变量 $\eta_1$ 和 $\xi_1$ 的测量模型中（现暂且忽略 $\eta_2$），则存在六个方程：

$$x_1 = \lambda_{11}\xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3$$
$$x_4 = \lambda_{41}\xi_1 + \delta_4$$
$$y_1 = \lambda_{11}\eta_1 + \epsilon_1$$
$$y_1 = \lambda_{21}\eta_1 + \epsilon_2$$

其中，$\lambda$ 是潜变量的系数或因子负荷，$\delta$ 和 $\varepsilon$ 分别是可观察的外生变量和内生变量的随机误差。因为所有的可观察变量都被假定以离差形式（即对于各自均值的离差）来测量，故方程中都没有截距项。这些方程中的元素可组合为向量或矩阵形式（每个潜变量的 $\lambda$ 都组成一列）。对于包括有 $\xi_1$ 的测量模型来说，我们有：

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

其中，矩阵 $\Lambda_x$ 只有一列，而向量 $\xi$ 只有一个元素，是因为我们只有一个 $\xi$ 变量。我们使用一般的矩阵代数符号，可以将测量方程表示为：

$$X = \Lambda_X\xi + \delta \tag{8.1}$$

其中，$\Lambda_x$ 包括了所有针对 x 的参数。类似地，我们以矩阵形式将与 y 有关的项目组合为：

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \end{bmatrix} \begin{bmatrix} \eta \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

或者，以更一般的矩阵代数符号表示为：

$$Y = \Lambda_y\eta + \epsilon \tag{8.2}$$

以相同的方式，我们可以很方便地将结构方程表示为：

$$\eta = B\eta + \Gamma\xi + \zeta \tag{8.3}$$

方程也可以重新表示为：

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

这一表达式将前面给出的涉及 $\eta_1$ 和 $\eta_2$ 的两个方程组合在一起了。

这里存在着四个方差——协方差矩阵：对于与可观察的外生变量相关的测量误差项来说有 $\Theta_\delta$；对于与可观察的内生变量的测量误差项来说有 $\Theta_\varepsilon$；对于外生的潜变量来说有 $\Phi$；对于内生的潜变量的误差项来说有 $\Psi$。在当前的

例子中：

$$\Phi = \begin{bmatrix} \sigma_{\xi 1}^2 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} \sigma_{\zeta 1}^2 & \\ 0 & \sigma_{\zeta 2}^2 \end{bmatrix}$$

$$\Theta_\delta = \begin{bmatrix} \sigma_{\delta 1}^2 & & & \\ 0 & \sigma_{\delta 2}^2 & & \\ 0 & 0 & \sigma_{\delta 3}^2 & \\ 0 & 0 & 0 & \sigma_{\delta 4}^2 \end{bmatrix}$$

$$\Theta_\epsilon = \begin{bmatrix} \sigma_{\epsilon 1}^2 & \\ 0 & \sigma_{\epsilon 2}^2 \end{bmatrix}$$

由上述可知，由于这里只有一个外生的潜变量，故在 $\Phi$ 中只有一个元素；并且所有的方差-协方差矩阵都是对称的。这意味着只有在对角线以下的元素才必须被写出。在当前的例子中，潜变量和可观察变量的所有误差项之间的协方差都被假定为零。

让我们用 NX 表示外生的可观察变量的数目；用 NY 表示内生的可观察变量的数目；用 NK 表示外生的潜变量的数目；用 NE 表示内生的潜变量的数目。四个方差-协方差矩阵的维度分别是：$\Phi$ 为 $NK \times NK$，$\Psi$ 为 $NE \times NE$，$\Theta_\delta$ 为 $NX \times NX$，$\Theta_\varepsilon$ 为 $NY \times NY$。前面定义过的四个参数矩阵的维度也可类似地表示，它们分别是：$\Lambda_x$ 为 $NX \times NK$，$\Lambda_y$ 为 $NY \times NE$，$B$ 为 $NE \times NE$，$\Gamma$ 为 $NE \times NK$。综合起来，在结构方程模型中，$B, \Gamma, \Lambda_x, \Lambda_y, \Phi, \Psi, \Theta_\delta, \Theta_\varepsilon$ 这八个矩阵中的零元素和非零元素的模式则可完全设定出一个模型。例如，对于不存在潜变量的经典的通径模型中，$\Lambda_x = I$，$\Lambda_y = I$，$\Theta_\delta = 0$，$\Theta_\varepsilon = 0$。同样，当 $B = 0$，$\Gamma = 0$，$\Psi = 0$，$\Lambda_y = 0$，$\Theta_\varepsilon = 0$ 时，一般模型就简化成为了一个验证性的因子分析（confirmatory factor analysis）。

许多结构方程模型都基于以下假设：

$$\Sigma = \Sigma(\theta) \tag{8.4}$$

其中，$\Sigma$ 是可观察变量的总体协方差矩阵，$\theta$ 是模型参数向量，$\Sigma(\theta)$ 是由模型推导出的协方差矩阵，或者说是以 $\theta$ 函数形式表达的协方差矩阵。换言之，结构方程模型的估计目标就是得到的参数将使模型假定下的方差-协方差矩阵 $\Sigma(\theta)$ 与观察到的方差-协方差矩阵 S 之间的差异最小化。

## 8.2.2 识别

在此，我们不将进行数学证明，它要用到矩阵 $\Sigma(\theta)$ 作为可观察的方差和协方差的函数，我们只是提出一些可以遵循的一般规则。在本节中，我们只处理递归模型（即非对称性关系的模型而不是外生变量间存在相关的模型）的

识别问题。对于非递归模型的识别问题,请参阅博伦(Bollen, 1989)。对于递归模型来说,它必须确保以下条件:

1. 在 NX + NY 个可观察变量中非冗余方差和协方差的总数大于八个矩阵 $\mathbf{B}$,$\mathbf{\Gamma}$,$\mathbf{\Lambda}_x$,$\mathbf{\Lambda}_y$,$\mathbf{\Phi}$,$\mathbf{\Psi}$,$\mathbf{\Theta}_\delta$,$\mathbf{\Theta}_\varepsilon$ 中的自由参数 K 的数目。

2. 所有的潜变量都必须被赋予一个测量单位。这可使用以下两种方式,并任选一种来获得:潜变量可以按其某一个指标变量来量纲化(scaled),或者也可以被标准化,从而有取值为 1 的单位方差。

3. 如果某一潜变量只有一个指标,那么,此潜变量就被假定为没有测量误差。

4. 研究者必须利用相关的识别规则来进一步确保模型的可识别性,如两步骤规则(two-step rule)、多指标多原因规则(multiple-indicator multiple-cause,以下简称 MIMIC)及沃尔德的秩规则(Bollen, 1989)。

## 8.2.3 估计

在一般结构方差模型中,可观察变量的方差-协方差矩阵 $\Sigma$ 为:

$$\mathbf{\Sigma} = E\left(\begin{bmatrix} Y \\ X \end{bmatrix}\begin{bmatrix} Y \\ X \end{bmatrix}'\right) = E\left(\begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}\right) \tag{8.5}$$

其中

$$\begin{bmatrix} Y \\ X \end{bmatrix}' = \begin{bmatrix} Y_1 Y_2 \cdots Y_{NY} X_1 X_2 \cdots X_{NX} \end{bmatrix}$$

将式(8.5)中最后一个矩阵的元素替换为潜变量和相关参数以及误差项,可以看出,一般的结构方程模型中的模型推导出的方差-协方差矩阵$\Sigma(\theta)$可以用八个矩阵表示为(Mueller, 1996:152):

$$\mathbf{\Sigma}(\theta) = \begin{bmatrix} \mathbf{\Lambda}_y(\mathbf{I} - \mathrm{B})^{-1}(\mathbf{\Gamma\Phi\Gamma}' + \mathbf{\Psi})\left[(\mathbf{I} - \mathrm{B})^{-1}\right]'\mathbf{\Lambda}_Y' + \mathbf{\Theta}_\epsilon & \mathbf{\Lambda}_y(\mathbf{I} - \mathrm{B})^{-1}\mathbf{\Gamma\Phi\Lambda}_X' \\ \mathbf{\Lambda}_x\mathbf{\Phi\Gamma}'\left[(\mathbf{I} - \mathrm{B})^{-1}\right]'\mathbf{\Lambda}_y & \mathbf{\Lambda}_x\mathbf{\Phi\Lambda}_X' + \mathbf{\Theta}_\delta \end{bmatrix} \tag{8.6}$$

从式(8.6)中,我们可以推导出对于通径模型或者验证性的因子模型的简化表达式,它们的八个矩阵中有一些是零矩阵或者单位矩阵。

对于一个可识别的结构方程模型进行估计的目标是得到估计值 $\hat{\mathbf{B}}$,$\hat{\mathbf{\Gamma}}$,$\hat{\mathbf{\Lambda}}_x$,$\hat{\mathbf{\Lambda}}_y$,$\hat{\mathbf{\Phi}}$,$\hat{\mathbf{\Psi}}$,$\hat{\mathbf{\Theta}}_\delta$,$\hat{\mathbf{\Theta}}_\varepsilon$,它们共同决定了由模型推导出的方差—协方差矩阵 $\mathbf{\Sigma}(\hat{\theta})$,以使 $\mathbf{\Sigma}(\hat{\theta})$能够近似 $\Sigma$。通过不断迭代,就可逐渐发现使得两者越来越近似的估计值。为了测量这种近似性,在每次迭代中都会对形式如 $F\left[S\mathbf{\Sigma}(\theta)\right]$ 的拟合函数进行评估,以确保 S($\Sigma$ 的样本估计)和 $\Sigma(\hat{\theta})$之间的距离不断缩小,直至达到最小。此时,拟合函数的值与前一次迭代的差距小于预

定标准。

最常用的两种估计方法是最大似然(ML)和广义最小二乘(GLS)。ML 和 GLS 都依赖于数据的多元正态分布假定。有时对于结构方程模型分析来说,这一假定可能太严格了。为了处理这一问题,诸如 LISREL 和 EQS 这样的结构方程模型软件实行一种称为“渐进性分布自由法”(the asymptotically distribution free method,以下简称 ADF)的更为灵活的估计技巧,此方法是以变量的四阶多元积矩为基础的(Browne, 1984)。

在诸如 ML 和 GLS 这样的较为严格的估计方法(要求多元正态分布假定)和 ADF 类型的方法(需要较少的假定)之间,还存在另一种基于椭圆分布理论(elliptical distribution theory)的估计方法。多元椭圆是一种分布类型,多元正态可视为其特例。我们应注意到,ML 和 GLS 估计方法在多元椭圆形式中和 ADF 形式中都有其对应。这些备选方案使得分析者在选择合适的估计方法时具有了更大的灵活性。

## 8.2.4 修正

初始假设的模型已经被设定、识别和估计了。假设模型的估计值已经收敛,且没有显示出估计中存在误差,数据分析者现在将要输出表格(或者面对电脑屏幕)。最佳的结果是假设完全被证实,初始假设的模型完美地拟合数据。但是,通常结果并不如此简单。假设的模型对数据的拟合可能从任何一种可用的拟合指标来看都相差甚远。研究者现在应立即回到画板前,从第一步重新开始。

在重新思考初始模型时,模型至少有四个方面可以加以完善:

1. 首先,初始模型是不是过于复杂了,某些参数可能是不必要的?通常研究者在刚开始时就把模型搞得过于复杂。某些参数可能无论从统计角度还是从实质角度来看都是没有存在依据的。

2. 是否有一些路径或参数本应该属于自由参数估计,却在初始模型中设定为固定参数?像 LISREL 软件提供修正指标,能在当前的固定参数中找到有潜在的估计价值的某些参数。

3. 初始模型的设定是否正确?这里的设定既包括参数和路径这些简单方面的设定,而且还包括其他可观察变量和潜变量的设定。

4. 由于线性回归传统的影响,通常那些初次进行结构方程模型的分析者会将那些代表固定分组的变量作为外生变量包括到模型中,如种族、性别、时间段或年龄组等。这些变量究竟是否可以被视为分组变量?换言之,对它们进行多组或多群体分析是否更合适些呢?

对于上述问题的回答将引导我们在下一次分析数据的尝试中对模型进行更为合理的设定。尤其是最后一点,它作为本节的重点今后还将更为详尽地加以讨论。

### 8.2.5 解释

一旦得到了一个令人满意的模型,最后的任务便是解释其结果。只要涉及的是直接效应,那么解释标准化和非标准化估计值的方式类似于线性回归,但是结构方程模型的复杂性给解释提出了更大的挑战。例如,除了社会经济地位对于学业成就产生的直接效应外,社会经济地位通过认知能力还将产生间接效应。因此,直接效应和间接效应的总和构成了它对于学术成就的总效应。

正式用数学语言来表述,矩阵 **B** 包括了模型中所有 η 变量之间的直接效应。它们的间接效应(IE)为:

$$\mathbf{IE}_{\eta\eta} = \mathrm{B}^2 + \mathrm{B}^3 + \mathrm{B}^4 + \cdots \tag{8.7}$$

其中,$\mathbf{B}^2$ 为矩阵乘积 **BB**,其他以此类推。一对 η 变量之间的中介变量的最大数目决定了式(8.7)中的幂应当是多大。换言之,任何一对 η 变量之间的通径的最多数目等于式(8.7)中的最高次幂。

总效应(TE)等于直接效应与间接效应之和,故其公式应为:

$$\mathbf{TE}_{\eta\eta} = \mathrm{B} + \mathrm{B}^2 + \mathrm{B}^3 + \mathrm{B}^4 + \cdots \tag{8.8}$$

同样,我们可以得到外生潜变量和内生潜变量之间的直接效应、间接效应和总效应。因为它们之间的直接效应包括在矩阵 **Γ** 中,我们必须将 **Γ** 与 **B** 相乘得到 IE

$$\mathbf{IE}_{\eta\xi} = (\mathrm{B} + \mathrm{B}^2 + \mathrm{B}^3 + \mathrm{B}^4 + \cdots)\mathbf{\Gamma} \tag{8.9}$$

为了获得其总效应,我们必须再一次加上 **Γ**,以代表直接效应,故得:

$$\mathbf{TE}_{\eta\xi} = (\mathbf{I} + \mathrm{B} + \mathrm{B}^2 + \mathrm{B}^3 + \mathrm{B}^4 + \cdots)\mathbf{\Gamma} \tag{8.10}$$

为了得到合理的解释,如同在经典的路径分析中一样,以上介绍的结构效应成分所采用的只是标准化参数。

## 8.3 均值与协方差结构

至此我们考察的还只是协方差结构。也就是说,我们忽略了潜变量的均值和测量方程的截距。在典型的结构方程模型应用中,结构方程的参数就是回归系数,变量的协方差结构包含着最重要的参数信息。在这样一种类型的模型中,可观察变量的均值并没有被分解为参数来进行估计。如果研究者感兴趣的只是斜率系数,无论是为了评估某一结构变量(潜变量)的效应,还是为了跨组比较斜率效应,这种模型都是可取的。在此种情形下,模型通过样本协方差矩阵 S 来加以设定、估计和检验。

但是,有时候数据分析者希望了解均值和截距。特别是在进行分组比较时,研究者可能对比较不同组中的同一潜变量的均值感兴趣。在此情形中,样

本均值向量$\left[\overline{Y}\ \overline{X}\right]'$也包含着重要的统计信息,样本协方差矩阵 S 和样本均值向量$\left[\overline{Y}\ \overline{X}\right]'$必须被同时加以分析。

为了有助于理解变量的均值与方程中的截距之间的关系,我们考察一个简单的双变量回归模型。与第 3 章中的符号一致,有一个解释变量的线性回归形式如下:

$$y = \beta_0 + \beta_1 x + e$$

其中,$\beta_0$ 是截距。虽然它一般情况下并不等于均值,但截距有助于确定 y 的均值。与通常情况一样,误差项的均值被假定为零。这样通过对两边取期望值,故得:

$$\mu_y = \beta_0^* + \beta_1 \mu_x$$

其中,$\mu_y$ 是 y 的均值,$\mu_x$ 是 x 的均值,$\beta_0^*$ 是新的截距,如果 x 变量以离差形式测量,则 x 的均值就是零,故 $\beta_0^*$ 就是 x 取均值时 y 的取值,且 $\beta_0^*$ 等于 $\beta_0$。在此表达式中,$\mu_y$ 由 $\beta_0^*$,$\beta_1$,$\mu_x$ 给出。这一表达式展示的方法被称为**结构均值模型**(structured means model),它将一个**均值结构**包容到协方差分析当中。

回想一下,在三个主要的 SEM 方程式(8.1)、式(8.2)、式(8.3)中,所有的随机变量都被假定均值为零。根据前面的讨论,现在我们放松这一假定,将式(8.1)、式(8.2)、式(8.3)扩展为除了前面八个参数矩阵外,还包括四个新的包括潜变量的截距和均值的参数矩阵。在 SEM 文献中符号表达存在不一致的地方,在下文中则采用 LISREL 的符号表达惯例。

与式(8.1)、式(8.2)、式(8.3)相对应,容纳了截距矩阵的三个基本方程变成了如下形式:

$$X = \tau_x + \Lambda_x \xi + \delta \tag{8.11}$$

$$Y = \tau_y + \Lambda_y \eta + \epsilon \tag{8.12}$$

$$\eta = \alpha + B\eta + \Gamma\xi + \zeta \tag{8.13}$$

其中,在理论上 $\tau_x$,$\tau_y$,$\alpha$ 等同于简单双变量回归中的截距项 $\beta_0$,虽然它们现在是维度分别等同于 Y(NY ×1),X(NX ×1),$\eta$(NE ×1)的包括了截距的向量。同样,我们假定误差项的期望值为零,因此,可以由此得到潜变量的均值。但我们需要一个最终的矩阵 $\kappa$ 来表示 $\xi$ 的均值 $E(\xi)$。由此,通过对式(8.13)取期望值,我们得到了 $\eta$ 的均值 $E(\eta)$ 为:

$$E(\eta) = (I - B)^{-1}(\alpha + \Gamma\kappa) \tag{8.14}$$

其中,$\kappa$ 的维度与 $\xi$ 相同为(NK ×1)。类似地,我们定义 X 和 Y 的均值为:

$$\mu_x = \tau_x + \Lambda_x \kappa \tag{8.15}$$

$$\mu_y = \tau_y \Lambda_y (I - B)^{-1}(\alpha + \Gamma\kappa) \tag{8.16}$$

在 LISREL 中,四个额外的矩阵 $\tau_x$,$\tau_y$,$\alpha$,$\kappa$ 在程序中被明确使用,并且被缺省设定为固定的零矩阵。这意味着,在缺省设定下,LISREL 估计的是一个协方差结构模型,除非这四个截距均值矩阵中有一些以不同方式设定。通过

自由设定这四个矩阵中的某些(或全部)元素,一个均值和协方差结构模型就可以得到估计。在 EQS 中,均值结构是通过引入一个固定“变量”V999 来解决的,并且需要对矩量矩阵(moment matrix)进行分析。除了协方差(或相关系数再加上标准差)外,最小的数据要求还包括可观察变量的均值向量。

总而言之,在一个简单的样本分析中,如果不设置进一步的条件,所有均值参数 $\tau_x$,$\tau_y$,$\alpha$,$\kappa$ 将是不可识别的。原则上讲,要估计的截距数目应当少于或者等于可观察变量的数目。在多组分析中(样本或总体)要比在单组分析中可以估计更多的均值,这一点将在下一节加以说明。

回想一下,对于估计协方差结构模型而言,一个基本条件是所有的潜变量都依照各自的一个指标而具有量纲集合。在相同的意义上,只有澄清潜变量均值的来源,它才能得到识别。让我们用 SES 的例子来说明。对于测量 SES 来说,如在本章前面所叙述的,它有四个指标。应用式(8.15)以量纲形式将其写出,则:

$$
\begin{aligned}
E(x_1) &= 0 + 1\kappa_1 \\
E(x_2) &= \tau_{x2} + \lambda_{21}\kappa_1 \\
E(x_3) &= \tau_{x3} + \lambda_{31}\kappa_1 \\
E(x_4) &= \tau_{x4} + \lambda_{41}\kappa_1
\end{aligned}
$$

其中,$\xi_1$(SES)的量纲被设定为是 $x_1$ 的量纲(如母亲的教育程度),故其他的 $\lambda$ 就因此得以识别。此外,$\xi_1$ 的零点被设定为与 $x_1$ 的零点相同,因此使得 $\kappa_1$ 等于 $E(x_1)$。故所有其他的 $\lambda$ 就得以识别,显然这种选择是主观的。我们完全可以将另一个 $\lambda$ 设定为 1,将另一个的 $\tau_x$ 设定为 0。只要外生的潜变量 SES($\xi_1$)的均值($\kappa_1$)得以识别,$\alpha$ 以及 $\eta$ 的均值也就可以识别了。

## 8.4　结构方程模型中的分组比较

与其他统计方法一样,不论组是固定的、随机选定的(如在实验设计中),还是非随机指定的(如在准实验设计中),在结构方程模型中的多组比较都是非常有用的。但是,由于结构方程模型具有相当的复杂性,例如,变量(包括可观察变量和潜变量)的潜在总数是多少,可观察变量和潜变量的关系如何,内生变量和外生变量的关系如何,有没有包括误差项相关的问题,故有相当多的途径进行组间比较。

首先,各组之间相等的约束条件可针对协方差结构模型中的八个矩阵 $\Lambda_x$,$\Lambda_y$,$B$,$\Gamma$,$\Phi$,$\Psi$,$\Theta_\delta$,$\Theta_\varepsilon$ 来加以检验。此外,还可检验均值结构中的四个矩阵 $\tau_x$,$\tau_y$,$\alpha$,$\kappa$ 在各组之间的参数相等假设。这样,我们建立分组比较的假设可以基于这十二个参数矩阵中某些或全部:

$$
\Lambda_{gx}, \Lambda_{gy}, B_g, \Gamma_g, \Phi_g, \Psi_g, \Theta_{g\delta}, \Theta_{g\epsilon}, \tau_{gx}, \tau_{gy}, \alpha_g, \kappa_g
$$

其中,下标“g”表示第 g 组,g = 1,2,…,G。虽然在结构方程模型的文献中下标和上标“g”都被使用,但为了与本书中表示组的符号相一致,这里使用的是下标“g”。如果研究者感兴趣的是要检验两组间所有结构关系的参数相等的假设,与广义线性模型中相似,零假设将是:

$$H_0: \quad B_g = B_h \wedge \Gamma_g = \Gamma_h$$

其中,$g \neq h$,“∧”的意思是“并且”,表示一个联合检验。

但是,结构方程模型的分布假设和模型设定的许多方面都表明,组间变异性的检验也应针对它们,而不应仅仅针对参数。这些方面包括可观察变量的多元分布、协方差矩阵 $\Sigma$、结构方程模型的形式及模型参数 $\theta$。下面我们将考察这四个问题。

## 8.4.1 多元分布的相同性

在结构方程模型中,最常见的分布假设就是多元正态分布。尽管它有非常严格的限定条件使许多来源的数据很容易违背这一假设,但是在大多数结构方程模型的应用中还是要假定分布是多元正态的,其原因在于它是正态分布的简单扩展,而正态分布是经典线性回归模型的基础。结构方程模型方法的最新发展已经使其他分布的假定成为可能,这已在前面的估计一节中有所讨论。多元椭圆分布和 ADF 方法可被视为是限定条件较少的分布假定。这些进展也使检验各组之间的分布是否相同成为可能。也就是说,是否某一组中的数据满足多元正态假定,而另一组的数据却没有满足呢?

为了回答此类问题,数据分析者必须对于可观察数据的多元分布提出一个假设。令 $Z = \begin{bmatrix} X & Y \end{bmatrix}$,并且用 md(Z)表示 Z 的多元分布。研究者可能对下列检验感兴趣:

$$H_0: \quad md(Z_1) = md(Z_2) = \cdots = md(Z_G)$$

即零假设陈述的是所有组都具有相同的多元分布。

对于这一假设并没有一个正规的检验。检验它的一个合理而略为不正规的办法是,分别对于每组数据拟合某一个结构方程模型。通过将所有组的模型设定保持一致,同时用基于多元正态分布、多元椭圆分布、ADF 的估计方法来拟合模型,研究者便可比较各个组在不同分布假定下的对同一组得到的模型拟合指标,如模型的 $\chi^2$ 或 BIC。在不同分布假定下对同一组的模型拟合的指标差异比另一组的同一差异大,则表明其背后的多元分布存在潜在的差异。我们如何断定,两组间的差异已经大到表明其多元分布不同的程度?一个可能成为决策的方法是将模型的 $\chi^2$ 转化为 BIC,再将 BIC 转化为贝叶斯因子。表 7.8 中介绍的基于后验几率比较模型的原则可以用来决定 g 组在两种分布假定下的不一致程度是否比 h 组更大,进而表明其背后的分布是不同的。

## 8.4.2 协方差矩阵的相同性

在使用结构方程模型进行分组比较时,另一个差异可能是不同组的协方

差矩阵的差异。这样，一个可能的假设就是所有的协方差矩阵都是相同的，即：

$$H_0: \quad \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_G$$

这个假设不是非常有用，因为我们只能观察到 $S_g$，而样本协方差矩阵的差异是可以归结为样本方差的差异。在此，我们真正想了解的是相关矩阵之间是否相同。协方差矩阵不同的原因有可能是因为在所有 G 组的变量的样本方差不同。但只要相关矩阵相同，结构关系的基础仍是相同的。

让我们用 $C_g$ 来表示 g 组的总体相关矩阵，它与 $\mathbf{\Sigma}_g$ 的关系为：

$$\mathbf{\Sigma}_g = \boldsymbol{\sigma}'_g C_g \boldsymbol{\sigma}_g$$

其中，$\boldsymbol{\sigma}_g$ 是 Z 中变量的标准差的对角矩阵。要进行检验的相等假设为：

$$H_0: \quad C_1 = C_2 = \cdots = C_G$$

为了检验这一假设，可将所有的 X 和 Y 变量视为 X 变量，然后分析输入协方差矩阵。对于每一个 x 变量来说，在测量方程中设定一个没有误差项的 ξ：

$$X = \mathbf{\Lambda}_x \xi$$

其中，估计的参数矩阵 $\mathbf{\Lambda}_x$ 等于 $\boldsymbol{\sigma}$，即非零对角元素正是变量的标准差。剩下要做的就是检验 $\xi_g$ 的相关矩阵之间是否相同，则

$$H_0: \quad \mathbf{\Phi}_1 = \mathbf{\Phi}_2 = \cdots = \mathbf{\Phi}_G$$

通过检验这一假设，我们就检验了相关矩阵是否相同，同时允许样本方差不同。但是，如果数据分析者认为各个组的方差应当相同，则 $\mathbf{\Lambda}_{gx}$（真正的标准差）可被约束为相同。这样我们检验的就是一个更加严格的假设，它要求各组之间的所有总体 $\mathbf{\Sigma}_g$ 都相同。

### 8.4.3　模型形式的相同性

模型形式上的某些差异可以通过约束参数为零并从模型中去除某些路径或者变量来加以设定。模型形式上的这些差异被认为是源于模型参数上的不同，这将在下一小节进行讨论。但是，有时模型形式上的差异并不轻易地被设定为对参数的约束。因此，我们将仿效博林采用过的方法（Bollen，1989），如果两个模型具有相同的参数矩阵，维度相同，固定参数、自由参数、约束参数的位置相同，则定义两个模型具有相同形式。根据该定义，我们将模型参数相同的问题局限于只有自由参数值被加以比较的情况。在大多数应用中，研究者假定两个模型的形式是相同的，他们关注的是在给定形式下参数取值的相同与否。

为了揭示两组间的不同形式，让我们假定感兴趣的概念是某人的社会交往强度，$x_1$，$x_2$，$x_3$，$x_4$ 分别代表了它与家庭、亲戚、朋友和邻居的交往。对于某一人群而言，设定的模型可能是：

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

而对于另一人群而言,模型可能还包括了第二个潜变量,将亲属社会交往与非亲属社会交往区分开来,这样就增加了矩阵 $\Lambda_x$ 和 $\xi$ 的维度:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ 0 & 1 \\ 0 & \lambda_{41} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

其中,$\xi_1$ 和 $\xi_2$ 之间的相关用参数 $\phi_{21}$ 表示。具有不同的形式的可能性可能会为来自不同总体的数据提供更好的拟合。

模型形式不同的另一种可能性是潜变量的层次不同。例如,在一个模型中只有一阶潜变量存在,而在另一个模型中二阶或者列高阶潜变量存在。通过在一般性的结构方程模型中使用社会交往强度作为内生的潜变量,第一个群体的模型只有一个一阶潜变量表示社会交往;而第二个群体的模型可能包括一个一般社会交往的二阶潜变量,它又可由来自亲属关系和非亲属关系的社会交往两个方面来代表。

检验模型形式是否相同是直截了当的。当模型具有不同形式时,通常它们是一个嵌套于另一个中的。LRT 和它的渐进等效方法都可以采用。但是,有时要比较的模型并不是一个嵌套于另一个中的。这时可采用贝叶斯模型比较方法,如基于贝叶斯因子的后验几率。

## 8.4.4 模型参数的相同性

假定模型形式相同,我们可以进一步检验模型参数是否相同。要考虑的假设可以分为三种类型:有关结构模型(概念模型)的假设;有关测量模型的假设;有关一般结构方程模型的假设。仿效博林提出的逐步增加检验的限定性的办法(Bollen, 1989),下面我们将介绍三个层次:

### 8.4.4.1 检验结构模型的相同性

通常研究者只对于结构参数感兴趣。假定不同组的模型形式在结构模型和测量模型上都相同,下一步自然是看结构参数是否在各组之间相同;如果它们相同,则结构误差项的协方差矩阵是否相同;最后,外生的潜变量之间的相关是否也相同:

$$H_{B\Gamma}: \quad B_g = B_h \wedge \Gamma_g = \Gamma_h$$

$$H_{B\Gamma\Psi}: \quad B_g = B_h \wedge \Gamma_g = \Gamma_h \wedge \Psi_g = \Psi_h$$

$$H_{B\Gamma\Psi\Phi}: \quad B_g = B_h \wedge \Gamma_g = \Gamma_h \wedge \Psi_g = \Psi_h \wedge \Phi_g = \Phi_h$$

其中，g 和 h 分别表示不同的组，每个假设都可用一些模型拟合统计量来评估。

### 8.4.4.2　检验测量模型的相同性

有时研究者可能对于测量参数也感兴趣。于是可以继续检验和考察跨组的测量参数是否相同；如果相同，可进一步考察测量误差项的协方差矩阵是否相同；最后，我们还可以考察测量参数、误差、外生潜变量的相关联合起来在各组间是否相同。

$$H_{\Lambda x}: \quad \boldsymbol{\Lambda}_{gx} = \boldsymbol{\Lambda}_{hx}$$

$$H_{\Lambda_x \Theta_\delta}: \quad \boldsymbol{\Lambda}_{gx} = \boldsymbol{\Lambda}_{hx} \wedge \boldsymbol{\Theta}_{g\delta} = \boldsymbol{\Theta}_{h\delta}$$

$$H_{\Lambda_x \Theta_\delta \Phi}: \quad \boldsymbol{\Lambda}_{gx} = \boldsymbol{\Lambda}_{hx} \wedge \boldsymbol{\Theta}_{g\delta} = \boldsymbol{\Theta}_{h\delta} \wedge \boldsymbol{\Phi}_g = \boldsymbol{\Phi}_h$$

在一般结构方程模型的某些应用中，研究者可能只对测量参数相同感兴趣，而对结构效应是否相同不感兴趣。这时上面给出的假设同样适用。

### 8.4.4.3　检验一般模型的相同性

最后，在一般结构方程模型中，研究者可能对测量模型和结构模型都感兴趣，因此，必须检验这两者的相同性。例如，研究者可能对下列相同性假设的检验感兴趣：

$$H_{B\Gamma}, H_{\Lambda}, H_{\Lambda B\Gamma}, H_{\Lambda B\Gamma\Psi}, H_{\Lambda B\Gamma\Theta}, H_{\Lambda B\Gamma\Psi\Theta}, H_{\Lambda B\Gamma\Psi\Theta\Phi}$$

其中，$\boldsymbol{\Lambda}$ 既代表 $\boldsymbol{\Lambda}_x$，也代表 $\boldsymbol{\Lambda}_y$；而 $\boldsymbol{\Theta}$ 既代表 $\boldsymbol{\Theta}_\delta$，也代表 $\boldsymbol{\Theta}_\varepsilon$。

由于经常需要检验此类相同性假设，所有的主要结构方程模型软件中都内含用来检验跨组的模型参数的相同性约束的函数，通常使用的是拉格朗日乘子检验（LMT）。这种用法将在示例一节中用 EQS 和 LISREL 来演示。

### 8.4.4.4　检验均值结构的相同性

上面的讨论主要集中在检验协方差结构的相同性上。同样，我们也可以检验均值结构之间的相同性，这只是要涉及专为四个截距—均值矩阵（intercept-mean matrices）建立假设。我们可以有：

$$H_{\tau_x\tau_y}: \quad \boldsymbol{\tau}_{gx} = \boldsymbol{\tau}_{hx} \wedge \boldsymbol{\tau}_{gy} = \boldsymbol{\tau}_{hy}$$

$$H_{\alpha\kappa}: \quad \boldsymbol{\alpha}_g = \boldsymbol{\alpha}_h \wedge \boldsymbol{\kappa}_g = \boldsymbol{\kappa}_h$$

$$H_{\tau_x\tau_y\alpha\kappa}: \quad \boldsymbol{\tau}_{gx} = \boldsymbol{\tau}_{hx} \wedge \boldsymbol{\tau}_{gy} = \boldsymbol{\tau}_{hy} \wedge \boldsymbol{\alpha}_g = \boldsymbol{\alpha}_h \wedge \boldsymbol{\kappa}_g = \boldsymbol{\kappa}_h$$

可以将上述进行组合来检验你想要的特定假设。某些研究者可能会对潜变量得分的均值特别感兴趣。前面我们已经讨论过必须用一个可观察的截距的量纲来作为潜变量均值的量纲。但在进行分组比较时这一点并不必要，正如在示例一节将要演示的一样，我们只要使用某一组中潜变量的均值来作为比较基准（起始点）。只要研究者愿意，均值结构相同假设可与协方差结构相同假设以多种不同方式加以组合。

# 8.5 示　例

在本节中,我们将使用结构方程模型来考察一个有两个组的例子。在美国,启蒙教育计划(Head Start)是服务于学习成绩差的孩子的一项著名的教育项目。桑博姆(Sörbom,1981)在启蒙教育夏季计划中重新分析了303个白人孩子的样本,其中包括参加了计划的样本($N_1=148$)和用以配对的对照样本($N_2=155$)。在这一节中,我们将比较这两个样本的相关矩阵、估计的分布假设、协方差结构和均值结构。比较它们的均值结构将使我们可以评估认知能力这一潜变量的得分均值,从而确定项目是否成功。但是,我们必须控制孩子的社会经济地位(SES)背景。由于认知能力和社会经济地位都用多项指标测量,因此,有必要采用结构方程模型。

如本章中前面所介绍的,社会经济地位可采用四个可观察变量来测量,认知能力可采用两个可观察变量来测量:

$x_{g1}$:母亲的教育(MaEd)

$x_{g2}$:父亲的教育(PaEd)

$x_{g3}$:父亲的职业(PaOcc)

$x_{g4}$:家庭收入(FamInc)

$y_{g1}$:都市阅读测验得分(MRT)

$y_{g2}$:伊利诺心理语言能力测验得分(ITPA)

下标"g"表示样本是属于启蒙教育计划中的成员($g=1$)还是属于对照组($g=2$)。在表8.1和表8.2中分别列出了启蒙教育计划组和对照组的六个指标。为了便于将数据复制到计算机程序中,我们特意将相关矩阵分列两个表,而没有在一个表中以分列对角线两边的形式列出。这样,两个矩阵便可以包括进入后面附录中列出的计算机程序中。同时,在表中分别列出了变量的均值和标准差。

**表8.1　启蒙教育计划组的均值、标准值和相关矩阵**

| 变量 | 母亲教育 (MaEd) | 父亲教育 (PaEd) | 父亲职业 (PaOcc) | 家庭收入 (FamInc) | MRT得分 (MRT) | ITPA得分 (MRT) |
|---|---|---|---|---|---|---|
| 母亲教育 | 1.000 | | | | | |
| 父亲教育 | 0.441 | 1.000 | | | | |
| 父亲职业 | 0.220 | 0.203 | 1.000 | | | |
| 家庭收入 | 0.304 | 0.182 | 0.377 | 1.000 | | |
| MRT得分 | 0.274 | 0.265 | 0.208 | 0.084 | 1.000 | |
| ITPA得分 | 0.270 | 0.122 | 0.251 | 0.198 | 0.664 | 1.000 |
| 均值 | 3.520 | 3.081 | 2.088 | 5.358 | 19.672 | 9.562 |
| 标准差 | 1.332 | 1.281 | 1.075 | 2.648 | 3.764 | 2.677 |

注:$N_1=148$。

表 8.2　对照组的均值、标准值和相关矩阵

| 变量 | 母亲教育 | 父亲教育 | 父亲职业 | 家庭收入 | MRT 得分 | ITPA 得分 |
|---|---|---|---|---|---|---|
| 母亲教育 | 1.000 | | | | | |
| 父亲教育 | 0.484 | 1.000 | | | | |
| 父亲职业 | 0.224 | 0.342 | 1.000 | | | |
| 家庭收入 | 0.268 | 0.215 | 0.387 | 1.000 | | |
| MRT 得分 | 0.230 | 0.215 | 0.196 | 0.115 | 1.000 | |
| ITPA 得分 | 0.265 | 0.297 | 0.234 | 0.162 | 0.635 | 1.000 |
| 均值 | 3.839 | 3.291 | 2.600 | 6.435 | 20.415 | 10.070 |
| 标准差 | 1.360 | 1.195 | 1.193 | 3.239 | 3.900 | 2.719 |

注：$N_2 = 155$。

### 8.5.1　比较相关矩阵

通常，在结构方程模型中开始检验参数是否相同之前，我们可能会对检验相关矩阵是否相同的假设感兴趣：

$$H_0: \quad C_1 = C_2$$

使用前一节介绍的方法，假设可以用本章结尾处 8. A. 1 节和 8. A. 2 节中的 EQS 程序和 LISREL 程序来进行检验。检验结果在表 8.3 中给出，从 EQS 程序的输出结果中得到了单个相关相等的拉格朗日乘子检验（LMT），从 EQS 和 LISREL 程序的输出结果中得到了整体检验。如图 8.1 所示为从 LISREL 8.12中得到的相关结构相等的示意图。由于其数目较多，图中的估计值和曲线显得过密。依据 LMT 来进行判断，两组间只有 $\phi_{52}$ 和 $\phi_{62}$ 存在显著差异。但从总体来看，两个相关矩阵彼此是看不出差异的，最大似然卡方值（ML $\chi^2$）为 10.543，自由度为 15（$p=0.784$）。ERLS $\chi^2$ 也给出了完全相同的结论。

表 8.3　相关矩阵相同的 LM 卡方值

| 变量 | 母亲教育 | 父亲教育 | 父亲职业 | 家庭收入 | MRT 得分 |
|---|---|---|---|---|---|
| 母亲教育 | | | | | |
| 父亲教育 | 0.107 | | | | |
| 父亲职业 | 0.109 | 1.475 | | | |
| 家庭收入 | 0.090 | 0.139 | 0.014 | | |
| MRT 得分 | 0.021 | 3.906 | 0.001 | 0.903 | |
| ITPA 得分 | 0.285 | 6.258 | 0.118 | 0.615 | 0.519 |
| ML 卡方值 | | | 10.543 | | |
| ERLS 卡方值 | | | 10.378 | | |
| 自由度 | | | 15 | | |

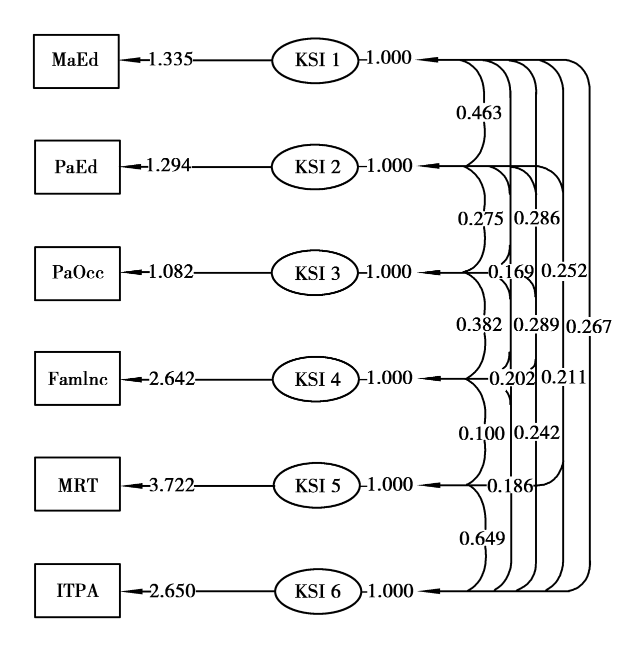



图 8.1　启蒙教育计划组和对照组相关矩阵相等

## 8.5.2　比较协方差结构和多元分布

为了比较两个协方差结构,首先我们必须设定测量模型和结构模型。我们用 $\xi_{g1}$ 表示社会经济地位(SES),用 $\eta_{g1}$ 表示认知能力。这样,可观察变量与潜变量之间的关系就可以下列测量方程来表示:

$$x_{g1} = \lambda_{g11}\xi_1 + \delta_{g1}$$
$$x_{g2} = \lambda_{g21}\xi_1 + \delta_{g2}$$
$$x_{g3} = \lambda_{g31}\xi_1 + \delta_{g3}$$
$$x_{g4} = \lambda_{g41}\xi_1 + \delta_{g4}$$
$$y_{g1} = \lambda_{g11}\eta_1 + \epsilon_{g1}$$
$$y_{g2} = \lambda_{g21}\eta_1 + \epsilon_{g2}$$

其中,加入下标"g"是为了表示属于哪一组。概念模型由下列结构方程定义:

$$\eta_{g1} = \gamma_{g11}\xi_{g1} + \zeta_{g1}$$

在结构方程中也加入了下标"g"。

在以上模型的设定下,我们估计了三个模型:第一个是对启蒙教育计划组,第二个是对对照组,第三个是通过两组的大多数参数设定为相同,对两个样本合并起来进行。其结果如表 8.4 所示,在 8.A.3 和 8.A.4 节中给出了第三个模型的计算机程序。图 8.2 是由 LISREL 8.12 给出的协方差结构相同模型的通径图。参数估计也包括在路径图中。注意,其中没有给出代表误差项的部分。

**表 8.4　启蒙教育数据的协方差结构模型的 ML 估计**

| 参数 | 启蒙教育计划组 估计值 | 启蒙教育计划组 标准误 | 对照组 估计值 | 对照组 标准误 | 两组合并 估计值 | 两组合并 标准误 | 两组合并 LM 卡方 |
|---|---|---|---|---|---|---|---|
| $\Lambda_x:\lambda_{21}$ | 0.669 | 0.203 | 1.033 | 0.235 | 0.874 | 0.152 | 0.857 |
| $\lambda_{31}$ | 0.963 | 0.272 | 1.259 | 0.362 | 1.109 | 0.224 | 0.100 |
| $\lambda_{41}$ | 2.399 | 0.678 | 2.793 | 0.819 | 2.603 | 0.530 | 0.001 |
| $\Lambda_y:\lambda_{21}$ | 0.876 | 0.214 | 0.885 | 0.219 | 0.862 | 0.150 | 0.005 |
| $\Gamma:\gamma_{21}$ | 1.950 | 0.729 | 2.101 | 0.796 | 2.049 | 0.542 | 0.080 |
| $\Theta_\delta:\theta_{21}$ | 0.469 | 0.149 | 0.405 | 0.140 | 0.426 | 0.102 | 0.127 |
| ML： | | | | | | | |
| $\chi^2$ | 15.655 | | 8.756 | | 26.017 | | |
| NFI | 0.917 | | 0.955 | | 0.932 | | |
| NNFI | 0.983 | | 0.979 | | 0.974 | | |
| ERLS： | | | | | | | |
| $\chi^2$ | 15.148 | | 8.572 | | 25.528 | | |
| NFI | 0.925 | | 0.962 | | 0.940 | | |
| NNFI | 0.907 | | 0.984 | | 0.979 | | |
| 自由度 | 7 | | 7 | | 20 | | |

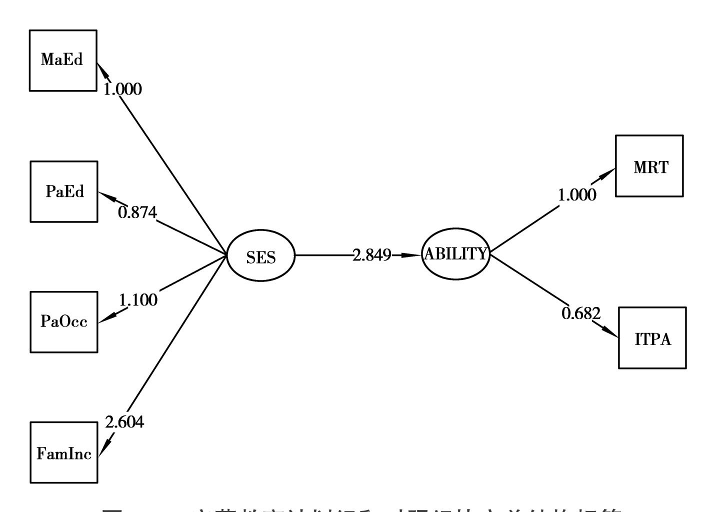



图 8.2　启蒙教育计划组和对照组协方差结构相等

首先，让我们分别考察 SEM 背后的对两组的多元正态假设。将多元正态假设放松为多元椭圆假设，启蒙教育计划组的卡方值提高了 0.5，而对照组的卡方值提高了 0.2。不同分布假设导致的卡方提高程度是微不足道的，这表明使用多元正态假设没有任何问题。

从两组的协方差结构分析的参数估计值中，可以看出一些差异，但似乎不存在严重的不一致。为了证实这种猜想，我们检验以下假设：

$$H_{\Lambda_1 \Gamma_1 \theta_{\delta 121}} = H_{\Lambda_2 \Gamma_2 \theta_{\delta 221}}$$

其中,由于我们对于检验测量方程中 x 变量的误差项方差是否相同并不感兴趣,故 $\theta_{\delta 121}$ 被单独挑出。矩阵 $\Lambda_x, \Lambda_y, \Gamma$ 中的参数被假定为在两组中相同。在相同性(约束)假定下的参数值大致位于启蒙教育计划组和对照组的参数估计之间。从单个的拉格朗日乘子检验统计量(LMT)或者从整体的似然比检验(LRT)来判断($\chi^2 = 26.017, df = 20$),两组模型参数间不存在差异。

## 8.5.3 比较均值和协方差结构

现在让我们把均值结构引入到比较中来。为此,我们需要在测量方程和结构方程中加入截距项:

$$\begin{aligned}
x_{g1} &= \tau_{g1} + \lambda_{g11}\xi_1 + \delta_{g1} \\
x_{g2} &= \tau_{g2} + \lambda_{g21}\xi_1 + \delta_{g2} \\
x_{g3} &= \tau_{g3} + \lambda_{g31}\xi_1 + \delta_{g3} \\
x_{g4} &= \tau_{g4} + \lambda_{g41}\xi_1 + \delta_{g4} \\
y_{g1} &= \tau_{g1} + \lambda_{g11}\eta_1 + \epsilon_{g1} \\
y_{g2} &= \tau_{g2} + \lambda_{g21}\eta_1 + \epsilon_{g2}
\end{aligned}$$

其中,$\tau_{g1}$ 是测量方程中解释第 g 组中变量 $x_{g1}$ 的截距项。结构方程也要进行相应修正:

$$\eta_{g1} = \alpha_{g1} + \gamma_{g11}\xi_{g1} + \zeta_{g1}$$

其中,$\alpha_{g1}$ 是 $\eta_{g1}$ 的截距项。注意,$\kappa$ 并未进入系统,除非同式(8.14)~式(8.16)中一样,方程以变量均值的形式表示。我们再一次估计三个均值和协方差模型,第一个模型对第一组估计,第二个模型对第二组估计,第三个模型则假设两组相同。

$$H_{\tau_1 \Lambda_1 \Gamma_1 \theta_{\delta 121}} = H_{\tau_2 \Lambda_2 \Gamma_2 \theta_{\delta 221}}$$

其中,加入 $\tau_g$ 以反映截距相同的假设。注意,矩阵 $\alpha_g$ 和 $\kappa_g$ 并未列入联合检验的假设中。为了拟合一个可识别模型,潜变量“社会经济地位”和“认知能力”的均值在对照组中被设定为零,这样在启蒙教育计划组中的对应值就可能以一个初始点来计量,或者以一个基准来估计。$\alpha$ 和 $\kappa$ 的相同与否可用参数估计直接评估。其结果如表 8.5 所示,在 8.A.5 和 8.A.6 节中将给出第三个模型的计算机程序。图 8.3 是由 LISREL 8.12 给出的均值协方差结构相同模型的通径图。由于这些参数包括在通径图中,因此,所得到的参数与前面只假定协方差结构相同的模型是不同的。

结果再一次说明,从两个分离模型中得到的参数估计彼此之间没有显著不同。第三个模型假定两组中的所有参数都相同,而它并没有对任何一个参数相同性约束给出显著的 LMT。整体的似然比检验(LRT)表明两组参数不存在差异($\chi^2 = 27.542, df = 22$)。但是,两组的 SES($\kappa$)水平非常不同,启蒙教育组中的水平显著低于对照组。只要控制了“社会经济地位”的效应,认知能力($\alpha$)的平均水平并不存在显著差异,这一结果对启蒙教育计划项目的有效性提出了疑问。

表 8.5　启蒙教育数据的均值和协方差结构模型的 ML 估计

| | 启蒙教育计划组 | | 对照组 | | 两组合并 | | |
|---|---|---|---|---|---|---|---|
| 参数 | 估计值 | 标准误 | 估计值 | 标准误 | 估计值 | 标准误 | LM 卡方 |
| $\tau_x:\tau_1$ | 3.520 | 0.110 | 3.839 | 0.110 | 3.870 | 0.094 | 0.033 |
| $\tau_2$ | 3.081 | 0.106 | 3.291 | 0.096 | 3.340 | 0.083 | 0.772 |
| $\tau_3$ | 2.088 | 0.089 | 2.600 | 0.096 | 2.574 | 0.090 | 0.582 |
| $\tau_4$ | 5.358 | 0.218 | 6.435 | 0.261 | 6.421 | 0.228 | 0.011 |
| $\Lambda_x:\lambda_{21}$ | 0.669 | 0.203 | 1.033 | 0.235 | 0.853 | 0.144 | 1.455 |
| $\lambda_{31}$ | 0.963 | 0.272 | 1.258 | 0.361 | 1.213 | 0.224 | 0.003 |
| $\lambda_{41}$ | 2.399 | 0.677 | 2.791 | 0.818 | 2.796 | 0.520 | 0.002 |
| $\tau_y:\tau_1$ | 19.672 | 0.310 | 20.415 | 0.314 | 20.358 | 0.287 | 0.193 |
| $\tau_2$ | 9.562 | 0.221 | 10.070 | 0.219 | 10.085 | 0.217 | 0.193 |
| $\Lambda_y:\lambda_{21}$ | 0.876 | 0.214 | 0.885 | 0.219 | 0.848 | 0.141 | 0.001 |
| $\kappa:\kappa_1$ | — | — | — | — | -0.381 | 0.103 | 13.683* |
| $\Gamma:\gamma_{21}$ | 1.950 | 0.729 | 2.101 | 0.795 | 2.146 | 0.554 | 0.108 |
| $\alpha:\alpha_1$ | — | — | — | — | 0.184 | 0.379 | 0.236* |
| $\Theta_\delta:\theta_{21}$ | 0.469 | 0.149 | 0.404 | 0.140 | 0.453 | 0.099 | 0.103 |
| ML： | | | | | | | |
| $\chi^2$ | 15.655 | | 8.756 | | 27.542 | | |
| NFI | 0.917 | | 0.955 | | 0.928 | | |
| NNFI | 0.893 | | 0.979 | | 0.987 | | |
| ERLS： | | | | | | | |
| $\chi^2$ | 15.148 | | 8.572 | | 26.858 | | |
| NFI | 0.994 | | 0.997 | | 0.995 | | |
| NNFI | 0.993 | | 0.999 | | 0.999 | | |
| 自由度 | 7 | | 7 | | 24 | | |

注：* 这些卡方值的计算方法是估计值与标准误的比值的平方，因为对照组的 κ 和 α 被设定为零。

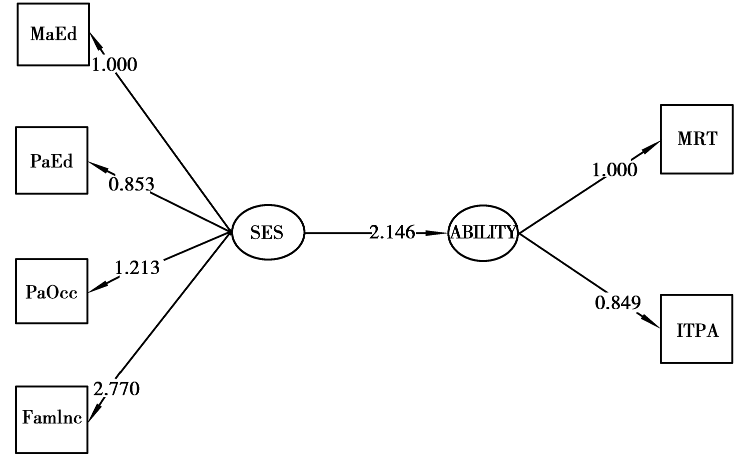



图 8.3　启蒙教育计划组和对照组的均值和协方差结构相等

# 8.A 计算机程序示例

为了节省篇幅,每个相关矩阵的中间四行被略去了。请参见表 8.1 和表 8.2 进行输入。

## 8.A.1 比较相关矩阵的 EQS 程序

```
/ title
  Head Start Two-Group Correlation Structure Model-the Head
  Start Group
/specifications
  case =148; variable =6; analysis =covariance; matrix =correlation;
  method =erls; groups =2;
/ labels
  v1 =maed; v2 =paed; v3 =paocc; v4 = faminc; v5 =MRT; v6 = ITPA;
  f1 =SES; f2 =cogable;
/equations
  v1 = * f1;
  v2 = * f2;
  v3 = * f3;
  v4 = * f4;
  v5 = * f5;
  v6 = * f6;
/ variances
  f1 to f6 =1;
/covariances
  f1 to f6 = * ;
/ matrix
  1.000
  …
  0.270 0.122 0.251 0.198 0.664 1.000
/standard deviations
 1.332 1.281 1.075 2.648 3.764 2.677
/end
/ title
  Head Start Two-Group Correlation Structure Model-the Control Group
/specifications
 case =155; variable =6; analysis =covariance; matrix =correlation;
```

```
  method=erls;
/labels
   v1=maed; v2=paed; v3=paocc; v4=faminc; v5=MRT; v6=ITPA;
   f1=SES; f2=cogable;
/equations
  v1 = * f1;
  v2 = * f2;
  v3 = * f3;
  v4 = * f4;
  v5 = * f5;
  v6 = * f6;
/variances
   f1 to f6 =1;
/covariances
  f1 to f6 = * ;
/matrix
   1.000
   …
   0.265 0.297 0.234 0.162 0.635 1.000
/standard deviations
  1.360 1.195 1.193 3.239 3.900 2.719
/constraints
  (1,f1,f2)=(2,f1,f2)
  (1,f1,f3)=(2,f1,f3)
  (1,f1,f4)=(2,f1,f4)
  (1,f1,f5)=(2,f1,f5)
  (1,f1,f6)=(2,f1,f6)
  (1,f2,f3)=(2,f2,f3)
  (1,f2,f4)=(2,f2,f4)
  (1,f2,f5)=(2,f2,f5)
  (1,f2,f6)=(2,f2,f6)
  (1,f3,f4)=(2,f3,f4)
  (1,f3,f5)=(2,f3,f5)
  (1,f3,f6)=(2,f3,f6)
  (1,f4,f5)=(2,f4,f5)
  (1,f4,f6)=(2,f4,f6)
  (1,f5,f6)=(2,f5,f6)
/lmtest
/end
```

## 8.A.2 比较相关矩阵的 LISREL 程序

```
Head Start Two-Group Correlation Structure Model-the Head Start Group
DA NO=148 NI=6 NG=2
KM*
  1.000
  …
  0.270 0.122 0.251 0.198 0.664 1.000
SD*
  1.332 1.281 1.075 2.648 3.764 2.677
LA*
  maed paed paocc faminc MRT ITPA
MO NX=6 NK=6 TD=FI
FR LX 1 1 LX 2 2 LX 3 3 LX 4 4 LX 5 5 LX 6 6
OU SE TV ND=3

Head Start Two-Group Correlation Structure Model-the Control Group
DA NO=155
KM*
  1.000
  …
  0.265 0.297 0.234 0.162 0.635 1.000
SD*
  1.360 1.195 1.193 3.239 3.900 2.719
LA*
  maed paed paocc faminc MRT ITPA
MO NX=6 NK=6 TD=FI PH=IN
FR LX 1 1 LX 2 2 LX 3 3 LX 4 4 LX 5 5 LX 6 6
Path Diagram
OU SE TV ND=3
```

## 8.A.3 比较协方差结构的 EQS 程序

```
/title
  Head Start Two-Group Covariance Structure Model-the Head
  Start Group
/specifications
  case=148; variable=6; analysis=covariance; matrix=correlation;
  method=erls; groups=2;
```

```
/ labels
   v1 =maed; v2 =paed; v3 =paocc; v4 = faminc; v5 =MRT; v6 = ITPA;
   f1 =SES; f2 =cogable;
/equations
  v1 = f1 + e1;
  v2 = * f1 + e2;
  v3 = * f1 + e3;
  v4 = * f1 + e4;
  v5 = f2 + e5;
  v6 = * f2 + e6;
  f2 = * f1 + d1;
/ variances
   e1 to e6 = * ;
   d1 = * ;
/covariances
  e1,e2 = * ;
/ print effect =yes;/matrix
   1.000
   …
   0.270 0.122 0.251 0.198 0.664 1.000
/standard deviations
  1.332 1.281 1.075 2.648 3.764 2.677
/end
/ title
   Head Start Two-Group Covariance Structure Model-the control Group
/specifications
  case =155; variable =6; analysis =covariance; matrix =correlation;
  method =erls;
/ labels
   v1 =maed; v2 =paed; v3 =paocc; v4 = faminc; v5 =MRT; v6 = ITPA;
   f1 =SES; f2 =cogable;
/equations
  v1 = f1 + e1;
  v2 = * f1 + e2;
  v3 = * f1 + e3;
  v4 = * f1 + e4;
  v5 = f2 + e5;
  v6 = * f2 + e6;
  f2 = * f1 + d1;
```

```
/ variances
   e1 to e6 = * ;
   d1 = * ;
/covariances
  e1,e2 = * ;
/ print
   effect =yes;
/ matrix
   1.000
   …
   0.265 0.297 0.234 0.162 0.635 1.000
/standard deviations
  1.360 1.195 1.193 3.239 3.900 2.719
/constraints
  (1,v2,f1) =(2,v2,f1)
  (1,v3,f1) =(2,v3,f1)
  (1,v4,f1) =(2,v4,f1)
  (1,v6,f2) =(2,v6,f2)
  (1,f2,f1) =(2,f2,f1)
  (1,e1,e2) =(2,e1,e2)
/ lmtest
/end
```

## 8.A.4 比较协方差结构的 LISREL 程序

```
Head Start Two-Group Covariance Structure Model-the Head Start Group
DA NO =148 NI =6 NG =2
KM *
  1.000
  …
  0.270 0.122 0.251 0.198 0.664 1.000
SD *
  1.332 1.281 1.075 2.648 3.764 2.677
LA *
  maed paed paocc faminc MRT ITPA
SE  5 6 1 2 3 4
MO  NX =4 NK =1 NY =2 NE =1 LX = FR LY = FR TD = SY
LK  SES
LE  ABILITY
FI  LX 1 1 LY 1 1
```

```
VA  1 LX 1 1 LY 1 1 FR TD 2 1
OU  SE TV ND=3
Head Start Two-Group Covariance Structure Model-the Control Group
DA NO=155
KM*
 1.000
 …
 0.265 0.297 0.234 0.162 0.635 1.000
SD*
 1.360 1.195 1.193 3.239 3.900 2.719
LA*
 maed paed paocc faminc MRT ITPA
SE  5 6 1 2 3 4
MO  NX=4 NK=1 NY=2 NE=1 LX=IN LY=IN GA=IN TD=SP
LK  SES
LE  ABILITY
EQ  TD(1,2,1)=TD(2,2,1)
Path  Diagram
OU  SE TV ND=3
```

## 8.A.5 比较均值和协方差结构的 EQS 程序

```
/title
  Head Start Two-Group Mean Covariance Structure Model-the Head
  Start Group
/specifications
 case=148; variable=6; analysis=moment; matrix=correlation;
 method=erls; groups=2;
/labels
  v1=maed;v2=paed;v3=paocc;v4=faminc;v5=MRT;v6=ITPA;
  f1=SES;f2=cogable;
/equations
 v1=3*v999+f1+e1;
 v2=3*v999+1*f1+e2;
 v3=2*v999+1*f1+e3;
 v4=5*v999+2*f1+e4;
 v5=20*v999+f2+e5;
 v6=10*v999+1*f2+e6;
 f1=*v999+d1;
```

```
  f2 = * v999 + 2 * f1 + d2;
/ variances
   e1 to e6 =1.5 * ;
   d1 to d2 =0.5 * ;
/covariances
  e1,e2 = * ;
/ print effect =yes;/ matrix
   1.000
   …
   0.270 0.122 0.251 0.198 0.664 1.000
/standard deviations
  1.332 1.281 1.075 2.648 3.764 2.677
/ means
   3.520 3.081 2.088 5.358 19.672 9.562
/end/ title
  Head Start Two-Group Mean Covariance Structure Model-the control Group
/specifications
  case =155; variable =6; analysis =moment; matrix =correlation;
  method =erls;
/ labels
   v1 =maed; v2 =paed; v3 =paocc; v4 = faminc; v5 =MRT; v6 = ITPA;
   f1 =SES; f2 =cogable;
/equations
  v1 =3 * v999 + f1 + e1;
  v2 =3 * v999 +1 * f1 + e2;
  v3 =2 * v999 +1 * f1 + e3;
  v4 =5 * v999 +2 * f1 + e4;
  v5 =20 * v999 + f2 + e5;
  v6 =10 * v999 +1 * f2 + e6;
  f1 =0v999 + d1;
  f2 =0v999 +2 * f1 + d2;
/ variances
   e1 to e6 =1.5 * ;
   d1 to d2 =0.5 * ;
/covariances e1,e2 = * ; / print effect =yes; / matrix
  1.000
  …
  0.265 0.297 0.234 0.162 0.635 1.000
/standard deviations
```

```
  1.360 1.195 1.193 3.239 3.900 2.719
/ means
  3.839 3.291 2.600 6.435 20.415 10.070
/constraints
  (1,v1,v999)=(2,v1,v999)
  (1,v2,v999)=(2,v2,v999)
  (1,v3,v999)=(2,v3,v999)
  (1,v4,v999)=(2,v4,v999)
  (1,v5,v999)=(2,v5,v999)
  (1,v6,v999)=(2,v6,v999)
  (1,v2,f1)=(2,v2,f1)
  (1,v3,f1)=(2,v3,f1)
  (1,v4,f1)=(2,v4,f1)
  (1,v6,f2)=(2,v6,f2)
  (1,f2,f1)=(2,f2,f1)
  (1,e1,e2)=(2,e1,e2)
/ lmtest
/end
```

## 8.A.6　比较均值和协方差结构的 LISREL 程序

```
Head Start Two-Group Mean Covariance Structure Model-the Head Start Group
DA NO=148 NI=6 NG=2
KM*
  1.000
  …
  0.270 0.122 0.251 0.198 0.664 1.000
SD*
  1.332 1.281 1.075 2.648 3.764 2.677
ME*
  3.520 3.081 2.088 5.358 19.672 9.562
LA*
  maed paed paocc faminc MRT ITPA
SE  5 6 1 2 3 4
MO  NX=4 NK=1 NY=2 NE=1 LX=FR LY=FR TX=FR TY=FR AL=FR KA=FR TD=SY
LK  SES
LE  ABILITY FI LX 1 1 LY 1 1
VA  1 LX 1 1 LY 1 1
FR  TD 2 1
OU  SE TV ND=3
```

```
Head Start Two-Group Mean Covariance Structure Model-the Control Group
DA NO=155
KM*
  1.000
  …
  0.265 0.297 0.234 0.162 0.635 1.000
SD*
  1.360 1.195 1.193 3.239 3.900 2.719
ME*
  3.839 3.291 2.600 6.435 20.415 10.070
LA*
  maed paed paocc faminc MRT ITPA
SE  5 6 1 2 3 4
MO  NX=4 NK=1 NY=2 NE=1 LX=IN LY=IN GA=IN TX=IN TY=IN AL=FI KA=FI
   TD=SP
LK  SES
LE  ABILITY
EQ  TD(1,2,1)=TD(2,2,1)
Path  Diagram
OU  SE TV ND=3
```

第 9 章

# 类别潜变量的比较

Comparison with Categorical Latent Variables

## 9.1 导 论

上一章讨论了结构方程模型中的统计分组比较。在结构方程模型中存在着潜变量时,它们被假定为是连续的,且通常还被假定为是服从多元正态分布的,对于可观察变量也有同样的假定。但是,正如观察变量可以是连续的也可以是类别的(categorical)一样,潜变量同样也可能是这两种类型之一。在此,当一个变量的测量尺度是数量的或定距的时候,我们定义它是连续的;当其测量尺度是定序或定类的时候,我们定义它是类别的。潜变量代表的概念既可以定义为连续的,也可以定义为类别的。例如,在第 8 章中的社会经济地位和认知能力被当做连续潜变量来进行研究。但是,若代表“专业技能”或“英雄行为”概念的潜变量最好被当做类别变量来进行分析。

在社会科学和行为科学中,由于许多概念都难于直接测量,指标或显变量常常被作为潜在概念的间接测量而被加以观察。结构方程模型的发展仅仅是潜结构模型这一更大主题的一个方面,潜结构模型是对于包含潜变量的统计模型的一般性名称。拉扎斯菲尔德和亨利(Lazarsfeld and Henry, 1968)总结和讨论了包含类别变量的潜结构模型的早期研究。在近期的众多研究中,巴塞洛缪和诺特(Bartholomew and Knott, 1999)以及弗蒙特(Vermunt, 1997)不仅较好地介绍了类别潜结构模型,而且还介绍了它们与连续潜变量模型之间的关系,以及其他相关的一些方法,如包含了潜变量的因果对数线性模型。

潜变量和观察变量都可以是连续的或类别的,因此,可根据变量的测量层次来对潜结构模型进行分类(Bartholomew and Knott, 1999)。对于潜变量和观察变量的测量层次进行交叉分类则产生一个 2×2 表(见表 9.1)。

表 9.1　潜结构模型的类型

| 观察变量 | 潜变量 | |
| --- | --- | --- |
| | 类别的 | 连续的 |
| 类别的 | 潜类模型 | 潜特质模型<br>因子分析模型 |
| 连续的 | 潜类模型<br>潜层序模型 | 因子分析模型 |

在因子分析中,作为结构方程模型的特例,假定在一系列连续指标背后存在着一个或多个连续潜变量。在以项目反应理论(item response theory)为基础的潜特质模型(latent trait model)中,一个连续潜变量是用一系列类别的观察变量来测量的。传统的结构方程模型中的测量模型(或因子模型)当观察变量是类别时,变量类型的组合也是如此。在潜层序模型(latent profile model)(和具有线性限定的潜类模型)中,潜变量是类别的,而可观察指标是连续的。最后,当潜变量和观察变量都是类别时,则得到典型的潜类模型(latent class model)。注意,虽然我们说因子分析是结构方程模型的一个特例,但应用统计方法的最新进展已经使得上述分界不再清晰。例如,可把潜类模型或者拉希模型(Rasch model)作为测量模型包容到一般结构方程的框架中以研究潜变量的因果效应。

在本章中,我们将主要关注表 9.1 中的第一列,即潜类模型和相关的类别潜变量模型。当然我们也会简要地讨论一下潜特质模型,因为它们也是基于潜在数据基础上的,并且常常与潜类模型一起加以讨论。在 9.2 节中,我们将介绍潜类模型(简称 LCM);在 9.3 节我们将讨论潜特质模型(简称 LTM)。在 9.4 节中,将考察用以处理连续观察变量的潜层序模型(简称 LPM)。在 9.5 节中,将考察包含类别潜变量的结构方程模型类型的因果模型,这是对 9.6 节中讨论存在类别潜变量时的统计比较的一个准备工作。最后,在 9.7 节,我们以示例来说明存在类别潜变量时的比较。由于大多数用于类别潜结构模型的软件是相当专业的,附录中列出了一些主要软件,并在第二个附录中列出了用于示例的一些相关程序。

## 9.2　潜类模型

在本节中,首先介绍潜类模型。我们假定现有数据有 K 个观察变量(显变量)$X_1$ 到 $X_K$,分别代表指标 $x_1$ 到 $x_k$,观察样本量为 N;每个显变量既可以是二分的,也可以是定序的。假定每个显变量 $x_k$ 的取值通过指标 $\xi$ 都与一个类别潜变量 $\Xi$ 关联。在此,我们用上标表示变量,用下标表示其指标或水平。

与第 8 章中一致，我们继续使用符号 ξ 和 x，尽管这里的潜变量是类别的。

经典潜类模型的基本公式为：

$$\pi_{x_1x_2\cdots x_K} = \sum_{\xi} \pi_{\xi x_1x_2\cdots x_K} \tag{9.1}$$

其中

$$\pi_{\xi x_1x_2\cdots x_K} = \pi_{\xi}\pi_{x_1x_2\cdots x_K|\xi} = \pi_{\xi}\pi_{x_1|\xi}\pi_{x_2|\xi}\cdots\pi_{x_K|\xi} \tag{9.2}$$

其中，$\pi_{\xi}$ 表示属于潜类 ξ 的概率。$\pi_{(x_1x_2\cdots x_k)}$ 表示联合分布 $\Xi X_1X_2\cdots X_K$ 中属于单元格（$\xi, x_1, x_2, \cdots, x_k$）中的概率；$\pi_{x_k|\xi}$ 是变量 $X_K$ 在给定属于潜类 ξ 时属于类别 $x_k$ 的条件概率 $P(X_K = x_k | \Xi = \xi)$。

我们再次使用前一章中的社会经济地位（SES）的例子来说明。假定 SES 是一个类别变量，分为高、中、低三级；母亲教育、父亲教育、父亲职业、家庭收入都以定序类别来测量，分别为 $X_1$ 到 $X_4$（见图 9.1）。根据式（9.2），四个显变量的某一组特定取值属于某一类 SES 的联合概率为：

$$\pi_{x_1x_2x_3x_4} = \pi_{\xi}\pi_{x_1|\xi}\pi_{x_2|\xi}\pi_{x_3|\xi}\pi_{x_4|\xi}$$

我们可以总结出潜类模型的三个特性：首先，潜类被假定是穷尽且互斥的。潜在的 SES 各类别的概率加总应当等于 1。其次，每个显变量的类别被假定是穷尽且互斥的。这意味着在每个 SES 类别之内，观察变量的各类别的概率加总应当等于 1。最后，也是最基本的，潜类模型假定局部独立性：

$$X_K | \Xi \sim \text{independent} \tag{9.3}$$

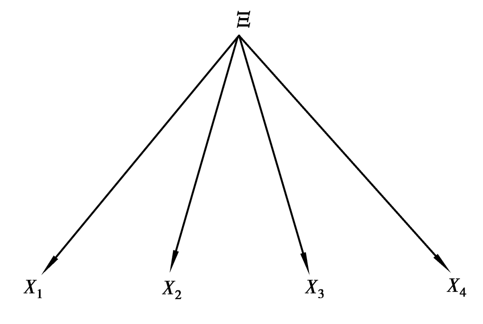



图 9.1　SES 的潜类模型

这一假定表明只要控制了潜变量，显变量彼此之间无关。也就是说，在每个 SES 类别内，观察变量彼此独立。

这就是经典的潜类模型，它一般按照拉扎斯菲尔德和亨利（Lazarsfeld and Henry, 1968）给出的式中（9.2）中概率形式参数化。但是哈伯曼（Haberman, 1979）证明式（9.2）中的经典潜类模型完全等同于层次对数线性模型 $\{\Xi X_1, \Xi X_2, \cdots, \Xi X_k\}$。这一对数线性模型可以写为：

$$\ln F_{\xi x_1x_2\cdots x_K} = u + u_{\xi}^{\Xi} + u_{x_1}^{X_1} + u_{x_2}^{X_2} + \cdots + u_{x_K}^{X_K} + u_{\xi x_1}^{\Xi X_1} + u_{\xi x_2}^{\Xi X_2} + \cdots + u_{\xi x_K}^{\Xi X_K} \tag{9.4}$$

式中　u——总均值效应（或截距）；

$u_{\xi}^{\Xi}$——潜变量 Ξ 的一阶效应；

$u_{x_k}^{X_K}$——显变量 $X_k$ 的一阶效应；

$u_{\xi x_k}^{\Xi X_K}$——Ξ 和 $X_k$ 的二阶效应或联合效应；

$F_{\xi x_1x_2\cdots x_k} = N\pi_{\xi x_1x_2\cdots x_k}$——单元格（$\xi, x_1, x_2, \cdots, x_k$）的期望频次。

其中：

$$\sum_{\xi} u_{\xi}^{\Xi} = \sum_{x_1} u_{x_1}^{X_1} = \cdots = \sum_{x_1} u_{x_K}^{X_K} = \sum_{\xi} u_{\xi x_1}^{\Xi X_1}$$
$$= \sum_{x_1} u_{\xi x_1}^{\Xi X_1} = \cdots = \sum_{\xi} u_{\xi x_K}^{\Xi X_K} = \sum_{x_K} u_{\xi x_1}^{\Xi X_K} = 0$$

通常的参数约束在此适用。对数线性的参数形式中只包括总均值、单变量项、潜变量 Ξ 与观察变量之间的双变量交互项。式(9.4)中不包含任何显变量之间的交互项,这是因为经典潜类模型假设中设定了显变量彼此之间是条件独立的。这就是重要的局部独立假定。

当然,对数线性传统和经典传统中的参数是相互关联的。根据哈伯曼(Haberman, 1979:551)和弗蒙特(Vermunt, 1997:51),式(9.2)中的条件概率可用式(9.4)中的对数线性参数项表达为:

$$\pi_{x_K \mid \xi} = \frac{\exp(u_{x_k}^{X_k} + u_{\xi x_k}^{\Xi X_k})}{\sum_{x_k} \exp(u_{x_k}^{X_k} + u_{\xi x_k}^{\Xi X_k})} \tag{9.5}$$

由于局部独立性假定,对每个可观察变量各自进行对数单位模型估计,与对具有层次边缘分布$\{\Xi X_1, \Xi X_2, \cdots, \Xi X_k\}$的列联表进行对数线性模型估计,两种估计方法中得到的潜变量与显变量之间的关系完全相同。

至此我们还只是讨论了未加限定的潜类模型。正如其他统计模型一样,我们可以对 LCM 加以限定。例如,在经典的参数化中,通常可以对潜概率和条件概率限定为某一固定值或限定为相等。另一方面,在对数线性的参数化中,通常可以对数线性参数加以限定,如对显变量效应、线性对线性关联、行效应、列效应和行列效应进行等效限定。非线性限定或不相等限定相对比较少见。对于条件响应概率加以特定的不相等限定,可以得到一种定序的潜类模型(Croon, 1990; Vermunt, 1997)。如果所有的显变量都是定序的,这种方法是相当有用的。

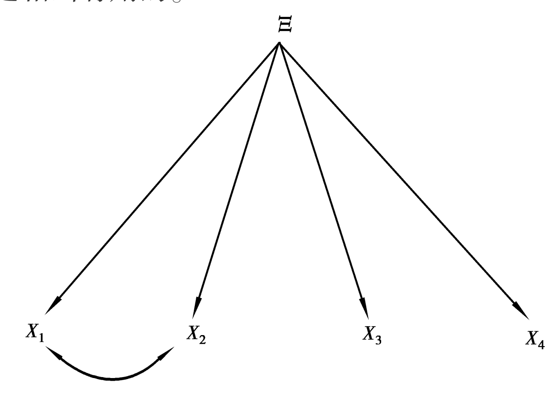



图 9.2 SES 的局部非独立模型

虽然局部独立性假定很重要,但是也可通过放松这一假定来扩展潜类模型。当控制潜变量后,某些指标可能彼此之间仍然相关。以"社会经济地位"为例,即使控制了潜变量"社会经济地位"的类别,母亲教育与父亲教育仍然是相关的。这就需要有局部非独立模型(local dependence model)。如图 9.2 所示说明它在"社会经济地位"一例中的应用。对数线性的潜类模型可以较容易地用来设定此类模型。哈格纳斯(Hagenaars, 1988)介绍了对局部非独立模型的设定。如在"社会经济地位"的一例中,对数线性模型就是$\{\Xi X_1, \Xi X_2, \Xi X_3, \Xi X_4, X_1 X_2\}$。也就是说,在母亲教育 $X_1$ 和父

亲教育 $X_2$ 之间存在着直接关联。在经典的参数化中,此模型也可写为:

$$\pi_{\xi x_1 x_2 x_3 x_4} = \pi_{\xi}\pi_{x_1 x_2 | \xi}\pi_{x_3 | \xi}\pi_{x_4 | \xi} \tag{9.6}$$

其中,$\pi_{x_1 x_2 | \xi}$通过一个没有三变量交互效应的对数线性模型来加以限定。这个模型等同于第 8 章中误差项存在相关的测量模型。

## 9.3　潜特质模型

在潜特质模型中,潜变量被称为潜特质(latent trait),它与外显反应之间的关系是通过对具有潜在得分 $\xi$ 的某一观察在项目 $X_K$ 的各个类别 $x_k$ 上的概率建模而得以分析。因此,关键是如何对条件概率 $\pi_{x_k|\xi}$进行建模。在项目反应理论中,潜特质被视为连续变量。

当观察与项目都位于同一单一维度的连续体时,概率 $\pi_{x_k|\xi}$随着差异 $\xi - x_k$ 而单调递增。使用不同的累积分布函数来对项目反应函数建模,这样便产生了两种不同的潜特质模型的类型:

一种被称为高斯型 S 形曲线或正态型 S 形曲线:

$$\pi_{x_k | \xi} = \Phi(\xi - x_k) \tag{9.7}$$

其中,$\Phi$ 表示累积标准正态分布函数。这样一个模型源于测量误差服从正态分布的假定。潜特质模型的高斯型可以使用允许类别变量的结构方程模型(SEM)软件加以估计。

另一种类型是逻辑斯蒂 S 曲线模型和拉希模型,它们对条件概率的建模如下:

$$\pi_{x_k | \xi} = \frac{\exp(\xi - x_k)}{1 + \exp(\xi - x_k)} \tag{9.8}$$

看起来,式(9.8)与潜类模型中的式(9.5)非常相似。其实两者之间的重要差异在于式(9.8)中潜变量是连续的,而潜类模型中潜变量是离散的。这一差异决定需考虑的函数是否服从累积逻辑斯蒂分布。项目反应理论的大多数文献都主要关注这一类型的潜特质模型。

以上的设定定义了一个具有连续特质的潜变量模型。在此基础上还可加以限定,以便将指标处理为定序变量,或将指标和潜变量之间的交互作用线性化。这就需要通过一个量纲参数 $a_k$ 来使得显变量的某些类别或潜变量的某些节点(node-points)具有固定值。具有此种限定的潜特质模型也被称为离散的潜特质模型。

当潜特质被离散化后,在潜特质模型的形式化表达中潜在序列中只有一个特定节点会起作用。博克(Bock, 1972)将潜特质模型扩展为:

$$\pi_{x_k | \xi} = \frac{\exp(a_k\xi + c_k)}{\sum_m \exp(a_k\xi + c_k)} \tag{9.9}$$

其中,$c_k = -a_k \times x_k$,将项目 $x_k$ 的各类加总。有如下假定,对于 $X_K$ 的 M 个反应类别中的 m 层来说,以下不等式成立:

$$a_{k1} < a_{k2} < \cdots a_{kM-1} < a_{kM} \tag{9.10}$$

此类模型可应用9.A 节中介绍的几个软件进行估计。但是,为了使潜特质的设定更为合理,最好具备较多的项目。潜特质模型最近已经被扩展为广义潜特质模型族,这一模型族可以分析服从各种指数族分布的显变量(Moustaki and Knott, 2000)。

## 9.4 连续指标的潜变量模型

潜类模型通过设定 $x_k$ 的条件分布的适当形式来处理二分或多分数据;而二项分布或多项分布则是合理的选择。如果观察变量是连续的,我们则考虑以下联合分布的分布形式应当如何(Bartholomew and Knott, 1999):

$$f(x_1, x_2, \cdots, x_K) = \sum_{\xi} \xi \prod_{k=1}^{K} h_k(x_k \mid \xi) \tag{9.11}$$

其中,$h_k(x_k \mid \xi)$是给定类别 $\xi$ 下 $x_k$ 的条件概率。服从此种设定的模型也被称为潜层序模型。

通常我们不完全了解 $h_k(x_k \mid \xi)$ 的形式。如果情况如此,很有必要考察以下边缘分布:

$$f(x_k) = \sum_{\xi} \xi h_k(x_k \mid \xi)$$

例如,偏态分布表明任何形式的正态分布组合都是不太可能的,而清晰的双峰分布则表明可能是两类正态分布的组合。拉扎斯菲尔德和亨利(Lazarsfeld and Henry, 1968)提出用矩法(the method of moments)来估计此类潜层序模型;最新的处理方法则依靠于期望最大化算法(EM method)。

如同在潜特质模型中使用线性限定一样,在为类别潜变量和数量型指标之间的关系建模时,对数线性模型也是最为灵活的。如海嫩(Heinen, 1993)指出的,尽管至今仍未被广泛应用,但包含了行关联或行关联的对数线性模型是可以处理此种情况的。尤其应当考虑如下情况:一系列二向交互表中所有的边缘都是由类别潜变量和数量型显变量组成的。显而易见,如果数量型可观察变量处在行上,我们应选择列效应关联的对数线性模型,同时对行加以线性限定;如果数量型可观察变量处在列上,我们应选择行效应关联的对数线性模型,同时对列加以线性限定。这一模型是限定模型,因为对于 $\Xi$ 的每个层次 $\xi$ 来说,其中的交互参数总数从 $X_K$ 的类别数减一减少为只有一个参数 $u_{\xi}^{\Xi X_k}(x_k - \bar{x}_k)$,其中,$x_k$ 是变量 $X_K$ 的数量取值。这样,式(9.4)中的二阶交互项则变为:

$$u_{\xi x_k}^{\Xi X_k} = u_{\xi}^{\Xi X_k}(x_k - \bar{x}_k) \tag{9.12}$$

其中,$k = 1, 2, \cdots, K$。这表明了对数线性形式是非常灵活的,通过简单地对对

数线性参数施加限定就可以产生潜类模型、潜特质模型和潜层序模型。

## 9.5　类别潜变量的因果模型

在本节中,我们将考察对潜变量模型的两种扩展。首先,我们允许在潜结构模型中存在多个潜变量。然后沿用古德曼(Goodman, 1973)适用于对数线性分析的修正后的路径模型,将其扩展到因果结构模型。此外,克洛格(Clogg, 1981)使用包含外部变量的潜类模型设定了适用于类别型数据的多指标多原因模型(MIMIC)。哈格纳斯(Hagenaars, 1993)较为通俗地介绍了包含潜变量的对数线性模型。

与因子分析模型可以包括多个因子一样,潜类模型也可以包括多个潜变量(Haberman, 1979)。此类模型可以通过对条件概率加以等值限定或者构建一个对数线性模型而加以设定。

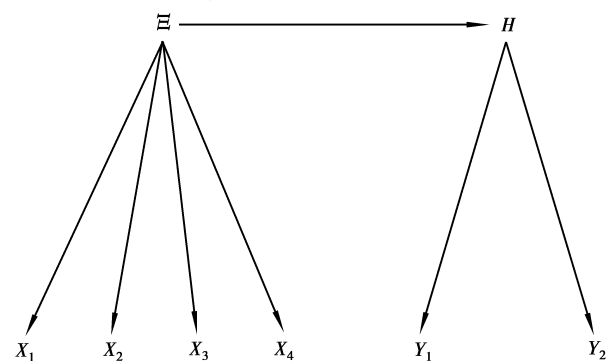



**图 9.3　对社会经济地位与认知能力中包括潜变量的因果模型**

我们继续使用社会经济地位的一例,但使用类别型变量。我们再次考察第 8 章中的启蒙教育计划的例子。$X_1$ 到 $X_4$ 是四个表示"SES"的变量,而 $Y_1$ 和 $Y_2$ 分别代表"MRT"和"ITPA",我们用 H 来代表潜变量"认知能力",其各个层次用 η 表示。其模型如图 9.3 所示。对前面结构方程模型检验过的模型加以修正,以适用于类别变量,这样便产生了对数线性潜类模型 $\{\Xi H, \Xi X_1, \Xi X_2, \Xi X_3, \Xi X_4, HY_1, HY_2\}$,其表达式为:

$$\ln F_{\xi\eta x_1x_2x_3x_4y_1y_2} = u + u_{\xi}^{\Xi} + u_{\eta}^{H} + u_{x_1}^{X_1} + u_{x_2}^{X_2} + u_{x_3}^{X_3} + u_{x_4}^{X_4} + u_{y_1}^{Y_1} + u_{y_1}^{Y_1} + u_{\xi\eta}^{\Xi H} + u_{\xi x_1}^{\Xi X_1} + u_{\xi x_2}^{\Xi X_2} + u_{\xi x_3}^{\Xi X_3} + u_{\xi x_4}^{\Xi X_4} + u_{\eta y_1}^{HY_1} + u_{\eta y_2}^{HY_2} \quad (9.13)$$

其中没有包含任何高于二阶的交互项。对这个对数线性模型的参数还可进一步加以限定,以便处理诸如连续性指标之类的问题。如果模型中有三个或更多的潜变量,只要将模型设定为不包含高于三阶或更高阶交互项即可。

这一模型还可用条件概率的形式表示:

$$\pi_{\xi x_1x_2x_3x_4\eta y_1y_2} = \pi_{\xi}\pi_{x_1|\xi}\pi_{x_2|\xi}\pi_{x_3|\xi}\pi_{x_4|\xi}\pi_{\eta|\xi}\pi_{y_1|\eta}\pi_{y_2|\eta} \quad (9.14)$$

其中包含了一种新型的条件概率 $\pi_{\eta|\xi}$,它只涉及潜变量。两种模型设定是完全等价的。

我们不难发现,它不同于结构方程模型,不论模型的设定是经典概率形式还是对数线性参数形式,其两个潜变量与第三个潜变量的交互作用很容易估计。例如,我们有两个外生潜变量 $\Xi_1$ 和 $\Xi_2$,一个内生潜变量 H,每个潜变量都有两个显变量表示。在对数线性回归模型中涉及潜变量的部分是 $\{\Xi_1\Xi_2, \Xi_1 H, \Xi_2 H, \Xi_1\Xi_2 H\}$。涉及显变量的分表(或边缘表)都被略去了。与此完全相同,条件概率形式中也将包括一项 $\pi_{\eta|\xi_1\xi_2}$。计算机程序 PANMARK, LCAT, LEM, Latent GOLD 都可以估计此类包含了类别潜变量的因果模型。LAT 和 NEWTON 也可以估计修正后的通径模型,不过要复杂些。

## 9.6 类别潜变量的比较

类似于结构方程模型中对于结构模型和测量模型的比较,研究者可以对比较各个组之间或各个总体之间的潜变量分布感兴趣,也可以对比较潜变量与显变量之间的关系感兴趣,也可以同时对两者都感兴趣。克洛格和古德曼(Clogg and Goodman, 1984, 1985)作出了重要的贡献,他们通过使用潜类模型的经典参数化建立了对多组的联立潜结构模型(simultaneous latent structure model)。他们的方法是在模型中加入一个分组变量,它可以影响到潜变量分布和条件概率。这种方法将在本节中做一介绍。

可以证明,对数线性模型是广义线性模型中的一种。假定其服从泊松分布,我们在第 6 章中已阐明了泊松族与指数族之间的关系。作为广义线性模型的一种,对数线性模型中的多组比较是比较容易于用标准方法处理的。需要进一步澄清的只是如何比较包含了类别潜变量的对数线性模型。虽然对数线性模型在一般情况下都是假定服从泊松分布的,但它也可以用多项分布或乘积多项分布作为分布基础。这三种抽样模型都被应用于对列联表的分析中。

**泊松分布**:我们对于观察总数 N 没有先验了解,故将表中每个单元格都视为在给定时间段内观察到的单个泊松过程。

**多项分布**:我们已知观察总数 N 为定值,然后根据列联表中变量的取值对每个观察进行交互分类。

**乘积多项分布**:对于行变量(或列变量、层次变量等)的每个类别,我们取一个样本量为 $n_{i+\cdots}$(或者 $n_{+i+\cdots}$等)的多项样本,然后对样本中的每个观察依据其在表中其他变量上的取值进行交互分类。

这样就会产生一个问题,即各个组背后的抽样模型是否完全相同。但由于这些抽样假导出的期望频次和对数线性模型拟合是相同的,故比较各个组的抽样模型并不是必要的。但是比较各个组之间在表中的实际抽样频次还是

非常必要的，是因为它们可能有不同的分布模式。

### 9.6.1　比较抽样分布

几乎所有的包含类别潜变量的模型都是基于列联表基础之上的。对数线性模型对于表中的频率分布很敏感。不同的组在同一个单元格内的频次可能会明显不同。一般的惯例是，列联表中的单元格最好包括至少有一个观察。如果一组在许多单元格中都抽到了零，而另一组则在同一些单元格中观察数很多时，就会出现问题。因此，在比较模型概率或参数之前，应先比较其抽样分布的模式，特别是比较抽到的零观察单元格。

零假设是假定各组抽到的零观察单元格具有相同模式：

$$H_0: \quad F_g^{(0)} = F_h^{(0)}$$

其中，上标“(0)”表示在某一模型下 g 组中抽到的期望频次为零的单元格的模式。为了检验这一假设，我们将抽到的零单元格视为结构上的零单元格，然后在零模型下分析各组数据，以比较各个组的零单元格（和非零单元格）。可以通过把各组两两配对，再比较它们的似然比统计量来进行似然比检验，但是对代表零单元格模式的单个对数线性参数进行检验会更容易一些。检验结果显著，则表明各个组的零单元格模型有差异。

### 9.6.2　比较变量间关联的类型和样式

在对数线性模型中，可以采用不同的方式对两向表的关联建模，如二阶交互、线性对线性关联、行效应关联、列效应关联及对数可积关联（log-multiplicative association）。同样，由一个潜变量与一个显变量构成的分表中的关联，或者由两个潜变量构成的分表中的关联，也可以在各个组中有不同的形式。为了进行这种比较，我们有如下原假设：

$$H_0: \quad A_{g\Xi X} = A_{h\Xi X} \wedge A_{gHY} = A_{hHY} \wedge A_{g\Xi H} = A_{h\Xi H}$$

其中，$A_{g\Xi X}$ 表示 g 组中 $\Xi$ 和 X 之间的关联类型，其他以此类推。比较的办法是：假设各组具有某种类型的关联，然后进行似然比检验。

有时，某一组中显变量之间的关联与另一组有着不同模式，结果导致各组中存在不同数量的潜变量。这样，零假设则变为：

$$H_0: \quad \dim(\Xi_g) = \dim(\Xi_h)$$

其中，$\dim(\Xi_g)$ 表示 g 组中潜变量 $\Xi$ 的维度。这一假设可以单独检验，也可与变量间关联类型的假设联合进行检验。在实践中这是一个非常有用的检验假设。

### 9.6.3　比较条件结构和响应概率

具有相同的基本数据模式（如抽样零单元格）的各组可能具有不同的潜结构。但是，具有不同的抽样零单元格模式的各组并不一定具有不同的潜结

构。变量之间的相关模式解释了大部分差异。如同在结构方程模型中比较模型参数一样，数据分析者可能对模型中的条件概率感兴趣，如给定潜变量下的显变量的条件概率、给定另一潜变量下的潜变量的条件概率、给定观察变量下的潜变量的条件概率等，这与在 MIMIC 模型中相同，即：

$H_{\pi_{x|\xi},\pi_{y|\eta},\pi_{\eta|\xi},\pi_{\eta|x}}$：

$$\pi_{g,x|\xi}=\pi_{h,x|\xi}\wedge\pi_{g,y|\eta}=\pi_{h,y|\eta}\wedge\pi_{g,\eta|\xi}=\pi_{h,\eta|\xi}\wedge\pi_{g,\eta|x}=\pi_{h,\eta|x}$$

其中，下标"g"用逗号与条件概率隔开来，以免把它和联合条件概率混淆。在此，如果我们使用结构方程模型中的术语，检验零假设 $H_{\pi_{x|\xi},\pi_{y|\eta}}$ 等同于检验测量模型相同；检验零假设 $H_{\pi_{\eta|\xi},\pi_{\eta|x}}$ 等同于检验结构模型相同。似然比检验可以很容易地适用于此类检验。特别要注意的是，上述零假设也可以用对数线性参数的形式表示，但用概率形式来表达更为简洁。在不同的软件中，研究者必须将上述等值检验表达为概率形式或者对数线性形式，但是不必同时使用两种形式。

## 9.6.4 比较潜分布和条件概率

潜类的分布比较有些类似于结构方程模型中的潜均值比较。研究者可能对比较潜类的构成而不是条件概率感兴趣。例如，启蒙教育计划组中的儿童与对照组的儿童在属于较低社会经济地位方面的概率是否有所差异；两组中的儿童在属于较高认知能力类别方面的概率是否有所差异。在这种研究情形下，要检验的零假设为：

$$H_{\pi_\xi,\pi_\eta}:\qquad \pi_{g,\xi}=\pi_{h,\xi}\wedge\pi_{g,\eta}=\pi_{h,\eta}$$

另一研究者可以同时对潜分布和条件概率都感兴趣。零假设就变为：

$H_{\pi_\xi,\pi_\eta,\pi_{x|\xi},\pi_{y|\eta},\pi_{\eta|\xi},\pi_{\eta|x}}$：

$$\pi_{g,\xi}=\pi_{h,\xi}\wedge\pi_{g,\eta}=\pi_{h,\eta}\wedge\pi_{g,x|\xi}=\pi_{h,x|\xi}\wedge\pi_{g,y|\eta}=\pi_{h,y|\eta}$$
$$\wedge\pi_{g,\eta|\xi}=\pi_{h,\eta|\xi}\wedge\pi_{g,\eta|x}=\pi_{h,\eta|x}$$

例如，对 SES 这一例同时进行潜分布和条件概率检验的潜类模型的形式为：

$$\pi_{g\xi x_1x_2x_3x_4}=\pi_g\pi_{\xi|g}\pi_{x_1|g\xi}\pi_{x_2|g\xi}\pi_{x_3|g\xi}\pi_{x_4|g\xi}\tag{9.15}$$

其中，下标"g"包含在所有的概率中。这一模型也可用对数线性术语描述为 $\{G\Xi X_1,G\Xi X_2,G\Xi X_3,G\Xi X_4\}$，其中，G 代表分组变量。与结构方程模型中的相同性检验一样，潜变量和条件概率的合理组合都可以构成一项研究假设。正如在示例一节将表明的，我们将再次使用似然比检验。

# 9.7 示　例

## 9.7.1 潜类分析中的比较

我们以没有外生变量的潜类模型开始示例。表 9.2 中关于美国天主教徒的数据是引自廖的文章(Liao, 1989b, 表 6.1),它是从 20 世纪 80 年代的综合社会调查项目(GSS)五年数据中得到的。

表 9.2　15 ~ 54 岁的白人女性天主教徒的宗教数据

| Relig16 | Reliten | Pray | Attend | 频次 |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 12 |
| 1 | 1 | 1 | 2 | 6 |
| 1 | 1 | 1 | 3 | 2 |
| 1 | 1 | 2 | 1 | 3 |
| 1 | 1 | 2 | 2 | 6 |
| 1 | 1 | 2 | 3 | 3 |
| 1 | 1 | 3 | 1 | 2 |
| 1 | 1 | 3 | 2 | 2 |
| 1 | 1 | 3 | 3 | 2 |
| 1 | 2 | 1 | 2 | 0 |
| 1 | 2 | 1 | 3 | 0 |
| 1 | 2 | 2 | 1 | 3 |
| 1 | 2 | 2 | 2 | 2 |
| 1 | 2 | 2 | 3 | 9 |
| 1 | 2 | 3 | 1 | 1 |
| 1 | 2 | 3 | 2 | 0 |
| 1 | 2 | 3 | 3 | 7 |
| 2 | 1 | 1 | 1 | 105 |
| 2 | 1 | 1 | 2 | 68 |
| 2 | 1 | 1 | 3 | 21 |
| 2 | 1 | 2 | 1 | 42 |
| 2 | 1 | 2 | 2 | 59 |
| 2 | 1 | 2 | 3 | 34 |
| 2 | 1 | 3 | 1 | 13 |
| 2 | 1 | 3 | 2 | 12 |
| 2 | 1 | 3 | 3 | 6 |
| 2 | 2 | 1 | 1 | 5 |
| 2 | 2 | 1 | 2 | 15 |

续表

| Relig16 | Reliten | Pray | Attend | 频次 |
|---|---|---|---|---|
| 2 | 2 | 1 | 3 | 42 |
| 2 | 2 | 2 | 1 | 5 |
| 2 | 2 | 2 | 2 | 19 |
| 2 | 2 | 2 | 3 | 104 |
| 2 | 2 | 3 | 1 | 3 |
| 2 | 2 | 3 | 2 | 11 |
| 2 | 2 | 3 | 3 | 51 |

注:“Relig16”代表回答者未成年时的宗教:1 = 其他宗教,2 = 天主教。

“Reliten”代表联系强度:1 = 不强,2 = 强。

“Pray”代表祷告频率:1 = 不到一天一次,2 = 一天一次,3 = 一天几次。

“Attend”代表参加宗教仪式的频率:1 = 一年一次或更少,2 = 多于一年一次但少于一周一次,3 = 一周一次或更多。

我们考虑的潜类模型中有三个显变量:Attend(参加仪式)、Pray(祷告)和 Reliten(联系强度)。Relig16 被当做一个分组变量。这些指标可更简化地表达为 $X_1$,$X_2$,$X_3$,G(为了避免在符号上的混乱,我们为分组变量单独使用一个符号),模型如图 9.4 所示,其中潜变量 Ξ 表示两类天主教成员。在拟合潜类模型之前,首先我们来考虑抽样零单元格的分布,在其他宗教中被教养的类别分表中存在着四个零单元格。为此,我们拟合一个对数线性模型来对比 Relig16 = 1 和 Relig16 = 2 两组的零单元格。我们检验如下假设:

$$H_0: \quad F_1^{(0)} = F_2^{(0)}$$

上述假设意味着两组宗教背景具有相同的零单元格模式。这可在 LEM 软件中通过具有效应编码的设计矩阵来完成,9. B. 1 给出了相应程序。

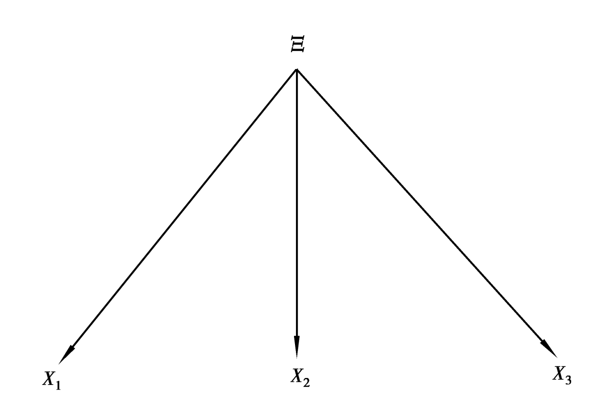



图 9.4　两组天主教徒的双类别潜变量模型

模型的整体拟合度 $L^2$ = 1 065,自由度 df = 34,但是我们对此并不感兴趣,

我们感兴趣的是比较代表两组的零单元格的对数线性参数。估计值为 −0.239，标准误为 0.081，沃德（wald）$\chi^2$ 值为 8.670，这表明两组在零单元格模式上具有显著差异。在此我们不必停止进一步的分析。但是，如果发现了显著差异，我们在进行更为精细的比较时必须牢记这一信息。

下一步，我们将考察潜变量 $\Xi$ 和显变量 $X_{gh}$ 之间关联的各种假设。在潜类模型中，两个变量一般都处理为类别变量，以二阶交互作用 $\{\Xi X_{gk}\}$ 相关。让我们用这一模型作为对两组进行模型比较的基准。9.B.2 节给出了相应的 LEM 程序。程序分别拟合了两组数据。既然 $X_1$ 和 $X_2$ 都有三个定序类别，我们接下来就考察边缘表 $\{\Xi X_1, \Xi X_2\}$ 的线性对线性（linear-by-linear）关联模型和列（$\Xi$）效应 RC-II 模型，9.B.3 节和 9.B.4 节分别给出了相应的 LEM 程序。表 9.3 中汇总了模型拟合信息。

**表 9.3　比较关联假设的类型**

| 关联 | −2£ | $\chi^2$ | $L^2$ | Cressie-Reed | 自由度 df |
|---|---|---|---|---|---|
| 第 1 组： | | | | | |
| 交互 | 295.747 | 6.405 | 7.496 | 6.504 | 6 |
| 线性对线性 | 299.119 | 11.297 | 10.869 | 10.434 | 8 |
| 列 RC-II | 299.119 | 11.302 | 10.869 | 10.440 | 8 |
| 第 2 组： | | | | | |
| 交互 | 3 085.393 | 13.887 | 14.294 | 13.998 | 6 |
| 线性对线性 | 3 090.108 | 18.860 | 19.009 | 18.919 | 8 |
| 列 RC-II | 3 090.108 | 18.960 | 19.009 | 18.919 | 8 |

观察一下各组中边缘表的二阶交互潜类模型和线性对线性关联（或列效应 RC-II 关联）潜类模型的 $\chi^2$ 值的差异。显而易见，对第一组来说交互作用的潜类模型较为合适，对第二组来说线性对线性或列效应 RC-II 关联的潜类模型较为合适。但是，在 0.05 的水平下这些差异都不显著，虽然在 0.10 的水平上是显著的。线性对线性模型和列效应 RC-II 关联模型之间的差别非常微小。因此，我们继续使用通常的包含二阶交互作用的潜类模型来比较概率。

最后，我们考察用经典的潜类概率和条件概率来比较两组。将考察四个潜类模型。在第一个模型中，所有的概率不论是否条件概率，都被认为在各组间相同，则：

$$\pi_{1,\xi} = \pi_{2,\xi} \wedge \pi_{1,x_k|\xi} = \pi_{2,x_k|\xi}$$

上述模型假定两组是完全同质的。这一模型的反面是完全异质性模型：

$$\pi_{1,\xi} \neq \pi_{2,\xi} \wedge \pi_{1,x_k|\xi} \neq \pi_{2,x_k|\xi}$$

在这两种极端情况之间，我们可以有部分同质性假定：

$$\pi_{1,\xi} \neq \pi_{2,\xi} \wedge \pi_{1,x_k|\xi} = \pi_{2,x_k|\xi}$$

以及部分异质性假定：

$$\pi_{1,\xi} \neq \pi_{2,\xi} \wedge \pi_{1,x_k|\xi} \neq \pi_{2,x_k|\xi} \wedge \pi_{1,\xi x_k} = \pi_{2,\xi x_k}$$

在9.B.5节到9.B.8节中列出了相应的LEM程序。我们将模型拟合结果汇总至表9.4中。

表9.4 比较潜类和条件概率

| 模型 | $-2\mathcal{L}$ | $\chi^2$ | $L^2$ | Cressie-Reed | 自由度df |
|---|---|---|---|---|---|
| 完全同质 | 3 791.947 | 25.487 | 27.652 | 27.652 | 23 |
| 完全异质 | 3 786.085 | 20.292 | 21.790 | 20.503 | 12 |
| 部分同质 | 3 791.871 | 25.516 | 27.576 | 25.614 | 22 |
| 部分异质 | 3 788.073 | 21.943 | 23.778 | 22.124 | 17 |

比较表9.4中几个模型的拟合指标就可以清楚地看出,与同质性模型相比较,完全异质性模型和部分异质性模型使用了更多的参数来容纳第二组,但是在模型拟合上却提高不多。两个同质性模型的拟合指标几乎是一样的,虽然Cressie-Reed统计量略为倾向于部分同质性模型。我们在表9.5中给出了部分同质性模型的概率估计。

因此,虽然不同宗教背景的两组之间条件响应概率被设为定值,但它们之间的潜分布仍然可以是不同。但是,如表9.5所示的这种差异不超过2%。这再一次表明:部分同质性和完全同质性模型之间是非常接近的。

表9.5 从部分同质性模型中得到的估计值

| 概率 | 估计值 | 标准误 |
|---|---|---|
| $\pi_g$ | | |
| 1 | 0.089 | 0.011 |
| 2 | 0.911 | 0.011 |
| $\pi_{\xi\mid g}$ | | |
| 1\|1 | 0.546 | 0.075 |
| 2\|1 | 0.454 | 0.075 |
| 1\|2 | 0.525 | 0.036 |
| 2\|2 | 0.475 | 0.036 |
| $\pi_{x_1\mid\xi}$ | | |
| 1\|1 | 0.510 | 0.035 |
| 2\|1 | 0.419 | 0.029 |
| 3\|1 | 0.071 | 0.033 |
| 1\|2 | 0.030 | 0.017 |
| 2\|2 | 0.160 | 0.027 |
| 3\|2 | 0.810 | 0.035 |
| $\pi_{x_2\mid\xi}$ | | |
| 1\|1 | 0.592 | 0.030 |
| 2\|1 | 0.327 | 0.028 |
| 3\|1 | 0.081 | 0.016 |
| 1\|2 | 0.214 | 0.028 |

续表

| 概率 | 估计值 | 标准误 |
| --- | --- | --- |
| 2\|2 | 0.532 | 0.030 |
| 3\|2 | 0.254 | 0.027 |
| $\pi_{x3\mid\xi}$ | | |
| 1\|1 | 0.965 | 0.026 |
| 2\|1 | 0.035 | 0.026 |
| 1\|2 | 0.171 | 0.044 |
| 2\|2 | 0.829 | 0.044 |

## 9.7.2 在包含类别潜变量的路径模型中的比较

在第二个例子中,我们将重新分析哈格纳斯的数据(Hagenaars, 1993, 表 5.1)。在他原本的表 5.2 中的模型 4 中,有着两个内生潜变量:以 $\eta_1$ 代表的 $H_1$(系统参与)和以 $\eta_2$ 代表的 $H_2$(抗议容忍)。潜变量 $H_1$ 有三个指标,A(系统反应程度),B(意识形态水平),E(传统参与),潜变量 $H_2$ 有两个指标,C(压抑潜能)和 D(抗议赞同),两个潜变量之间被假定有着因果关系。每个潜变量都被假定有着两种类别。有两个外部变量 T(教育)和 G(年龄)被假定对两个潜变量都存在因果关系。第三个外部变量 S(性别)在本例中被用作分组变量。这样,哈格纳斯的模型 4 就被我们加以改进,用来检验潜分布和模型参数在各个组之间的相同性,其模型如图 9.5 所示。在附录 9. B 的 9. B. 9 中给出了频数表。

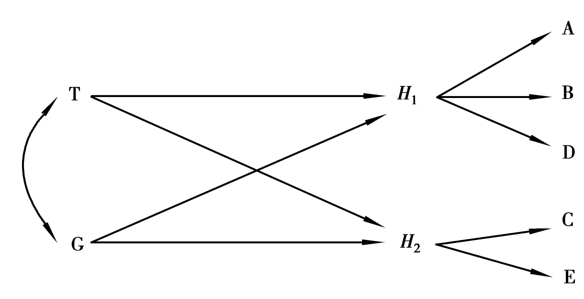



**图 9.5 不同性别的系统参与和抗议容忍的潜路径模型**

我们的第一个假设是联系两个类别潜变量与它们的类别指标的所有测量参数相同:

$$\boldsymbol{\pi}_{1,\mathbf{y}\mid\boldsymbol{\eta}} = \boldsymbol{\pi}_{2,\mathbf{y}\mid\boldsymbol{\eta}}$$

上述假设针对两个性别组中的所有 $\eta$ 及它们的指标(y)。黑体符号分别表示潜变量和观察变量的向量。

接下来,我们检验结构参数的相同性约束,用结构方程模型的术语来讲,是检验各个组的 $\beta$ 和 $\gamma$ 是否不同。这一检验的原假设为:

$$\pi_{1,\eta_2|\eta_1} = \pi_{2,\eta_2|\eta_1} \wedge \pi_{1,\eta_1|tg} = \pi_{2,\eta_1|tg} \wedge \pi_{1,\eta_2|tg} = \pi_{2,\eta_2|tg}$$

为了保持示例中原来的符号不变,上式中的 T 和 G 是用 t 和 g 表示的两个外部变量,它们并不代表分组。

最后,研究者可能想检验全部的模型参数和潜分布在两组间是否都相同:

$$\pi_{1,y|\eta} = \pi_{2,y|\eta} \wedge \pi_{1,\eta_2|\eta_1} = \pi_{2,\eta_2|\eta_1}$$
$$\wedge \pi_{1,\eta_1|tg} = \pi_{2,\eta_1|tg} \wedge \pi_{1,\eta_2|tg} = \pi_{2,\eta_2|tg} \wedge \pi_{1,\eta} = \pi_{2,\eta}$$

其中,黑体 $\boldsymbol{\eta}$ 表示 $\eta_1$ 和 $\eta_2$。

为了检验这三个假设,我们使用似然比检验来比较各组的 $-2£$ 或其他的 $\chi^2$ 类型统计量。在本例中,我们使用完全异质性模型参数作为比较基准。除了这个模型外,我们还需要三个模型来完成比较——结构参数(或概率)同质模型、测量参数(或概率)同质模型、包含测量参数和结构参数两者的模型参数(或概率)同质模型。9.B.10 到 9.B.13 给出了这四个模型的计算机程序。

在这四个模型中,只有完全异质性模型较好地拟合了数据,之后是假定结构参数同质但测量参数异质的模型。我们主要关注的是模型之间的比较,表 9.6 给出了各个模型的拟合优度统计量。

似然比统计量为 82.266,自由度为 24。假定不同性别的两组间的结构参数和测量参数都完全相同的模型明显劣于完全异质性模型,由此可以拒绝结构参数和测量参数的联合同质性假设。测量参数同质性假设在 0.01 水平上也被拒绝(LRT: 28.032, $df=10$)。结构参数同质性的假设在 0.05 水平上被拒绝,但是在 0.01 水平上被保留(LRT: 20.463, $df=9$)。总而言之,在两性间,所有模型参数(概率)都存在显著差异。

**表 9.6 比较四个包含了类别潜变量的通径模型**

| 模型 | $-2£$ | $\chi^2$ | $L^2$ | Cressie-Reed | 自由度 df |
|---|---|---|---|---|---|
| 完全同质 | 12 431.527 | 384.578 | 388.687 | 366.620 | 334 |
| 完全异质 | 12 459.559 | 411.510 | 416.719 | 393.540 | 344 |
| 部分同质 | 12 451.990 | 398.294 | 409.150 | 382.370 | 343 |
| 部分异质 | 12 513.793 | 465.904 | 470.953 | 446.882 | 358 |

在哈格纳斯的原文中(Hagenaars, 1993),性别是被当成了外部变量。我们在不同性别的两组间发现模型参数存在着显著差异,因此,最好进行分层分析。虽然对数线性的方法可以灵活地容纳交互项,但是使用分层模型更为容易,在概念上也更为清晰,这是因为两组间不仅结构参数而且测量参数都存在显著的差异。

## 9.A 针对类别潜变量的软件

有许多用于类别潜变量模型的专业软件。这里提供的并不是完整的软件

清单,而只包括笔者知晓并且在一定程度上熟悉的软件。此外,我们在此还将给大家介绍一些统计通用软件的相关功能,如 S-plus 中的一项功能和 SAS 中针对 LCMs 的一项宏命令。

### 9. A. 1　MLLSA

MLLSA(最大似然潜结构分析)最早是由克利福德· C· 克洛格(Clifford C. Clogg)为 IBM 大型计算机而编写。明尼苏达大学的斯科特· 伊莱亚森为 PC 机(DOS)编写了更新版本(MLLSA4. 0),并且将它包括到称为"类别数据分析系统"(CDAS)的软件中。软件 CDAS 包括了类别数据分析的许多程序,如层次和非层次对数线性模型、关联模型、潜类模型、清洗方法等。从斯科特· 伊莱亚森的主页 http://www. soc. umn. edu/ ~ eliason/CDAS. htm 上可以下载到 CDAS 或 MLLSA。

MLLSA 曾经是最常用的软件,但现在已出现了更为灵活和功能更强的其他软件。它非常易于使用,所使用的期望最大化(EM)算法确保了较快的收敛速度。

### 9. A. 2　LAT

LAT 是与 MLLSA 同时产生的,它由西北大学的谢尔比· 哈伯曼(Shelby Haberman)编写,它对于潜类模型进行了再参数化,采用的估计方法是牛顿—拉菲森算法(Newton-Raphson algorithm)。与其后继者 NEWTON 和 DNEWTON 一样,LAT 需要就其设计矩阵构建问题。这一过程对于某些人来说是麻烦的,而对于另一些人来说意味着灵活。如果需要进一步的信息,请咨询谢尔比· 哈伯曼(shelby@ gibbs. stats. nwu. edu)。

### 9. A. 3　PANMARK

PANMARK 由弗兰克· 范德普尔(Frank van de Pol)、罗尔夫· 兰根海茵(Rolf Langeheine)、W· 德· 荣格(W. de Jong)开发编写。它是基于期望最大化算法的一种非常灵活的 PC DOS 程序,可适用于类别潜结构模型和马尔科夫链模型。此程序可以在网页 http://ourworld. compuserve. com/homepages/jsuebersax/soft. htm[①] 上获取,并通过 ProGAMMA 公司销售。这一程序非常便捷,由于它使用多组起始值的特点可以避免出现局部最大值(Local maxima),用户估计较大模型时更为顺畅。

### 9. A. 4　LCAG

LCAG 由蒂尔伯格大学的雅克· 哈格纳斯(Jacgues Hagenaars)编写,它是

① 原书上的网址为:http://members. nbci. com/jsuebersax/soft. htm. 可能已变更。——译者注

一种可以进行局部非独立模型估计的非常灵活的程序。此程序的一个主要优点在于用户可以设定包含潜变量的对数线性模型。程序可以从作者那里得到(jacques. a. hagenaars@ kub. nl)。

## 9. A. 5 LEM

LEM 是由蒂尔伯格大学的杰罗恩· 弗蒙特(Jeroen Vermunt)编写的一项优秀的 PC 程序。它使用非常灵活,可以估计的模型种类非常多,包括用潜类、潜特质、局部非独立、事件史、关联、以对数线性形式表达的因果通径模型等。它还可以估计具有各种缺失机制的模型,可以检验参数约束。它的输出结果既包括对数线性估计,也包括经典的潜类概率估计。程序有 DOS 版本和 Windows 95 版本,使用非常便捷,而且还提供了非常好的示例。程序可从下列网址下载:http://www. uvt. nl/faculteiten/fsw/organisatie/departementen/mto/software2. html[①]。

## 9. A. 6 Latent GOLD

2000 年 9 月,“统计创新”公司发布了 Latent GOLD,它是由杰伊· 马吉德森(Jay Magidson)和杰罗恩· 弗蒙特合作两年的成果。这一软件脱胎于优秀的 LEM,包含了 LEM 中许多灵活的功能。某中一项卓越的功能允许用户在估计中应用贝叶斯常数。这些常数为概率参数估计设定一个先验分布,接近边界 0 和 1 的值将被认为是不太可能出现的,这样就为局部最大问题提供了一种新的解决方法。软件的试用版可在下列网址下载:http://latentclass. com。

Latent GOLD 与其他大部分软件还有一个重要的区别,就是它的 Windows 功能非常强大,可以与 SPSS 相媲美。事实上,它可以直接读取 SPSS 数据和 ASCII 数据。

## 9. A. 7 Mplus

在 20 世纪 80 年代后期,本特· 穆森(Bengt Muth´en’s)的 LISCOMP 曾是唯一可以处理类别数据的软件。近年来,他和林达· 穆森(Linda Muth´en’s)将 LISCOMP 的功能加以扩展,成为一种更为普遍的软件,它可以处理潜分布的各种组合,允许同时包括类别变量和连续变量。程序还可以处理多层次分析。更多的信息请参见网址:http://www. statmodel. com。

## 9. A. 8 LATCLASS, TWOMISS, 以及 POLYMISS

这些程序是为巴塞洛缪和诺特(Bartholomew and Knott, 1999)关于潜变

---

① 原书上的网址为:http://www. kub. nl/foculteiten/fsw/organistatie/departementen/mto/softwarez. html. 可能已变更。——译者注

量的著作服务的软件,可以在网址 http://www.arnoldpublishers.com/support/lvmfa2.htm 上免费下载。

LATCLASS 使用期望最大化(EM)算法来拟合标准的潜类模型(类别最多为 15 个,项目最多为 30 个)。TWOMISS 对于二分类数据使用对数单位—概率单位反应函数来拟合单因子或双因子潜特质模型,观察可以有缺失。POLYMISS 对于多分类数据使用对数单位—概率单位反应函数来拟合单因子或双因子潜特质模型,观察可以有缺失。这些程序都是基于 DOS 基础的(当然它们也可在 Windows 下运行),没有其他一些软件那么花哨,但是却非常实用。

### 9.A.9　lca.S 和 lcreg.sas

S-plus 中的 lca 功能(包含在文件 lca.S 中)可对编码为 0-1 的二分类数据进行经典的潜类分析。SAS 的宏命令 lcr(包含在文件 lcreg.sas)可以执行潜类回归(即多因多果模型)。两个文件可从网址 http://iosun01.biostat.jhsph.edu/~bjohnson/software/lvar/lvar.html 上下载。但是,SAS 的宏命令同样局限于编码为 0-1 的二分类响应数据,但协变量可以是数量型的。

## 9.B　示例中的计算机程序

### 9.B.1　比较样本零单元格的 LEM 程序

```
* comparing sampling zeros among multiple Groups
* Data:Liao(1989b,Table6.1)
*
* G is the group variable, relig16
* A,B and C are indicators of attend, pray, and reliten
* A design matrix is used to test equality of the zero cells
man 4
dim 2 2 3 3
lab G C B A
mod {cov(GABC,1)}
des [0 0 0 0 0 0 0 0 0
     1 1 0 1 0 0 0 1 0
     0 0 0 0 0 0 0 0 0
   -1 -1 0 -1 0 0 0 -1 0]
dat [12 6 2 3 6 3 2 2 2
       0 0 3 0 2 9 1 0 7
```

```
  105 68 21 42 59 34 13 12 6
     5 15 42  5 19 104 3 11 51]
```

### 9.B.2 比较二阶交互作用假定下的潜类模型的LEM程序

```
*Multiple group latent class analysis
*Data:Liao(1989b,Table6.1).
*
*G is the group variable, relig16;
*A,B and C are indicators of attend, pray, and reliten.
*Comparing marginal models:{AX} and {BX} are assumed
*fully interactive.

Lat  1
man  3
dim  2 2 3 3
lab  X C B A
mod  X A|X B|X C|X
dat  [12 6 2 3 6 3 2 2 2
       0 0 3 0 2 9 1 0 7]

*dat [105 68 21 42 59 34 13 12 6
*        5 15 42  5 19 104 3 11 51]
```

此程序运行两次:第一次完全如上所示;第二次则只保留第二组的数据,将第一组的数据加上星号,不再运行。

### 9.B.3 比较线性对线性关联假定下的潜类模型的LEM程序

```
*Multiple group latent class analysis
*Data:Liao(1989b,Table6.1).
*
*G is the group variable, relig16;
*A,B and C are indicators of attend, pray, and reliten.
*Comparing marginal models:{AX} and {BX} association are
*assumed linear-by-linear.

Lat  1
man  3
dim  2 2 3 3
lab  X C B A
mod  X A|X {A,spe(AX,1b)} B|X {B,spe(BX,1b)} C|X
```

```
dat   [12 6 2 3 6 3 2 2 2
        0 0 3 0 2 9 1 0 7]
*dat [105 68 21 42 59 34 13 12 6
*        5 15 42  5 19 104 3 11 51]
```

## 9. B. 4　比较列效应 RC-II 关联假定下的潜类模型的 LEM 程序

```
*  Multiple group latent class analysis
*  Data:Liao(1989b,Table6.1).
*
*  G is the group variable, relig16
*  A,B and C are indicators of attend, pray, and reliten.
*  Comparing marginal models:{AX} and {BX} association are
*  assumed RC-II.
Lat  1
man  3
dim  2 2 3 3
lab  X C B A
mod  X A|X {A,ass2(A,X,4a)} B|X {B,ass2(B,X,4a)} C|X
dat   [12 6 2 3 6 3 2 2 2
        0 0 3 0 2 9 1 0 7]
*dat [105 68 21 42 59 34 13 12 6
*        5 15 42  5 19 104 3 11 51]
```

## 9. B. 5　比较完全同质性假定下的潜类模型的 LEM 程序

```
*  Multiple group latent class analysis.
*  Data:Liao(1989b,Table6.1).
*
*  G is the group variable, relig16;
*  A,B and C are indicators of attend, pray, and reliten.
*  Completely homogeneous model: No variable depends on G.
Lat  1
man  4
dim  2 2 2 3 3
lab  X G C B A
mod  X A|X B|X C|X
dat   [12 6 2 3 6 3 2 2 2
        0 0 3 0 2 9 1 0 7
```

```
105 68 21 42 59 34 13 12 6
  5 15 42  5 19 104 3 11 51]
```

## 9.B.6 比较完全异质性假定下的潜类模型的 LEM 程序

```
*  Multiple group latent class analysis.
*  Data:Liao(1989b,Table6.1).
*
*  G is the group variable, relig16;
*  A,B and C are indicators of attend, pray, and reliten.
*  Completely homogeneous model: X, A|X,B|X and C|X all
*  depends on G

Lat  1
man  4
dim  2 2 2 3 3
lab  X G C B A
mod  G X|G A|XG B|XG C|XG
dat [12 6 2 3 6 3 2 2 2
     0 0 3 0 2 9 1 0 7

    105 68 21 42 59 34 13 12 6
      5 15 42  5 19 104  3 11 51]
```

## 9.B.7 比较部分异质性假定下的潜类模型的 LEM 程序

```
*  Multiple group latent class analysis.
*  Data:Liao(1989b,Table6.1).
*
*  G is the group variable, relig16;
*  A,B and C are indicators of attend, pray, and reliten.
*  partially homogeneous model: X as well as A, B and C
*  depends on G, but AX, BX, and CX do not.

Lat  1
man  4
dim  2 2 2 3 3
lab  X G C B A

mod G
    X|G
    A|XG {AX,AG}
    B|XG {BX,BG}
```

```
    C|XG {CX,CG}
dat [12 6 2 3 6 3 2 2 2
      0 0 3 0 2 9 1 0 7

    105 68 21 42 59 34 13 12 6
      5 15 42  5 19 104  3 11 51]
```

## 9.B.8 比较部分同质性假定下的潜类模型的 LEM 程序

```
*  Multiple group latent class analysis.
*  Data:Liao(1989b,Table6.1).
*
*  G is the group variable, relig16;
*  A,B and C are indicators of attend, pray, and reliten.
*  partially homogeneous model: X depends on G,
*  but AX, BX, and CX do not.
Lat  1
man  4
dim  2 2 2 3 3
lab  X G C B A
mod G X|G A|X B|X C|X
dat [12 6 2 3 6 3 2 2 2
      0 0 3 0 2 9 1 0 7

    105 68 21 42 59 34 13 12 6
      5 15 42  5 19 104  3 11 51]
```

## 9.B.9 例 2 中的数据

```
hageng3.fre:
 0   0   0   0   0   1   0   0   1   0   1   0
 0   1   0   0   1   0   0   0   0   0   1   0
 4   0   0   1   0   1   1   0   0   2   1   0
 6   3   1   3   1   1   4   4   1   1   1   0
 0   0   1   0   2   1   0   0   0   0   0   0
 2   3   0   0   2   2   0   3   2   0   4   1
 0   1   1   0   1   1   2   1   0   0   0   0
 4   7   4   0   4   2   4   2   0   0   3   2
 0   1   0   8   1   2   1   0   0   5   3   7
 0   1   0   3   5   3   0   0   1   1   2   2
 2   0   1   9   4   2   8   3   0  12   6   1
```

```
10   3   1   6   2   1   7   3   1  11   6   3
 0   1   0   2  10  13   3   2   2   8  21  47
 0   3   1   1  11  10   2   1   5   4  13  17
 1   1   2   9   7   7   3   3   0   7  11   8
 1   4   0   3   7   5   1   9   1   3   6   4
 0   0   0   1   0   1   0   0   0   1   0   0
 1   3   2   0   0   2   0   3   0   0   0   1
 2   1   0   0   0   0   2   1   0   1   0   1
11   6   5   2   4   0  13   7   4   1   1   1
 1   1   0   0   0   1   1   0   0   0   0   3
 1   5  10   1   0   5   0   2   4   1   4   5
 3   2   0   0   1   0   2   0   2   0   0   0
10  17   4   0   3   1   5   7   4   3   5   4
 0   0   0   3   0   1   0   2   0   4   2   4
 1   3   0   0   2   1   1   1   3   2   0   2
 2   1   0   7   0   3   6   0   1   8   4   1
13   7   1  10   4   3   9   8   6   4   6   9
 0   1   0   3   4   6   4   0   6   6   4  15
 5   7   4   2   7   8   2   6   9   3  15  24
 6   1   1   3   4   3   3   4   4   7   5   5
14   8   6   6   9   5   7   6   8   6  17   4
```

此数据原本是作为一个 LEM 例子使用的,来自于哈格纳斯(Hagenaars, 1993, 表 5.1)。数据是由下列八个变量交互分类而成的频次表:

A = 系统反应程度(1:低;2:高)

B = 意识形态水平(1:意识形态的;2:非意识形态的)

C = 压抑潜能(1:低;2:高)

D = 抗议赞同(1:低;2:高)

E = 传统参与(1:低;2:高)

S = 性别(1:男性;2:女性)

T = 教育(1:大学;2:大学以下)

G = 年龄(1:16 ~ 34;2:35 ~ 57;3:58 ~ 91)

频次数据保存于名为 hagen93.fre 的 ASCII 码文件中,在下面将会从 LEM 程序中调用。最右列的变量水平变化最快,最左列的变量水平变化最慢。

## 9.B.10 例 2 中在组间异质假定下的 LEM 程序

```
*  Loglinear Path Model with Latent Variables,
*  with All parameters Different between the Groups
*  Modified Model 4 of Table 5.2 using Sex as Group
```

```
*  Data: Hagenaars(1993, Table5.1)
*  A = System responsiveness
*  B = Ideological level
*  C = Repression potential
*  D = Protest approval
*  E = Conventional participation
*  S = Sex
*  T = Education
*  G = Age
lat  2
man  8
dim  2 2 2 2 2 2 2 2 2 3
lab  Y Z A B C D E S T G
mod  S
     GT|S {SGT}
     Y|SGT {STY,SGY}
     Z|SGTY {STZ,SGZ,SYZ}
     A|SY
     B|SY
     C|SZ
     D|SZ
     E|SY
dat  hagen93.fre
```

## 9.B.11 例 2 中在测量参数同质假定下的 LEM 程序

```
*  Loglinear Path Model with Latent Variables,
*  with Equal Measurement Parameters between the Groups
*  Modified Model 4 of Table 5.2 using Sex as Group
*  Data: Hagenaars(1993, Table5.1)
*  A = System responsiveness
*  B = Ideological level
*  C = Repression potential
*  D = Protest approval
*  E = Conventional participation
*  S = Sex
*  T = Education
*  G = Age
lat  2
```

```
man  8
dim  2 2 2 2 2 2 2 2 2 3
lab  Y Z A B C D E S T G
mod  S
     GT|S {SGT}
     Y|SGT {STY,SGY}
     Z|SGTY {STZ,SGZ,SYZ}
     A|Y
     B|Y
     C|Z
     D|Z
     E|Y
dat  hagen93.fre
```

## 9.B.12　例 2 中在结构参数同质假定下的 LEM 程序

```
*  Loglinear Path Model with Latent Variables,
*  with Equal Structural Parameters between the Groups
*  Modified Model 4 of Table 5.2 using Sex as Group
*  Data: Hagenaars(1993, Table5.1)
*  A = System responsiveness
*  B = Ideological level
*  C = Repression potential
*  D = Protest approval
*  E = Conventional participation
*  S = Sex
*  T = Education
*  G = Age

lat  2
man  8
dim  2 2 2 2 2 2 2 2 2 3
lab  Y Z A B C D E S T G
mod  S
     GT|S {SGT}
     Y|GT {TY,GY}
     Z|GTY {TZ,GZ,YZ}
     A|SY
     B|SY
     C|SZ
```

```
    D|SZ
    E|SY
dat  hagen93.fre
```

## 9.B.13　例 2 中在模型参数同质假定下的 LEM 程序

```
*  Loglinear Path Model with Latent Variables,
*  with Equal Constraints for All Model Parameters
*  and Latent Distributions between the Groups
*  Modified Model 4 of Table 5.2 using Sex as Group
*  Data: Hagenaars(1993, Table5.1)
*  A = System responsiveness
*  B = Ideological level
*  C = Repression potential
*  D = Protest approval
*  E = Conventional participation
*  S = Sex
*  T = Education
*  G = Age
lat  2
man  8
dim  2 2 2 2 2 2 2 2 2 3
lab  Y Z A B C D E S T G
mod
   GT {GT}
   Y|GT {TY,GY}
   Z|GTY {TZ,GZ,YZ}
   A|Y
   B|Y
   C|Z
   D|Z
   E|Y
dat hagen93.fre
```

第 10 章

# 多水平分析中的比较

Comparison in Multilevel Analysis

## 10.1 导　论

在第 3 ~9 章中,我们考察了“分组”作为外部变量时的统计比较。它们被视为某一固定变量的类别,某一特定模型(如广义线性模型或者潜变量模型)在各类别间被加以比较。有时并不是整个模型在各个组之间被假定为相同,而是某一个次级模型(submodel),如测量模型或结构模型被假定为相同。因此,模型相同或等值可能有着多种含意。多数情况下,我们仅关注模型参数,但研究者同样对分布假定和模型结构相同的假设检验感兴趣。

在本章中,我们将考察另一种可能性。我们将分组变量视为统计模型的内在部分,而非外部变量。这一区分主要是概念层面的,而非实际层面的;因为在前面章节中,作为一种建立固定效应模型的方法,我们已使用了包含分组变量的交互项来检验相同性假设。在概念层面上,“组”可被视为个体观察从中被抽取和被测量的单位。如果情况如此,则“组”就构成了在多水平分析中的一个“水平”。

一般将“组”引入多水平分析中可采用两种方式。在跨国研究中,个人是从一些国家中被抽取的,每个国家的特性都是独特而且已知的。在这种情形下,分组效应和组间差别最好在模型中设定为固定的。这一过程就是目前为止我们处理分组的方法。“组”同样可以是从一个较大总体中随机抽取出来的,其成员可能有一些与其他组不同的独特特性,也可能没有。这就需要一个包含随机效应的模型。

本章我们将介绍对于层次数据的随机效应模型,它是广义线性模型的一种主要扩展形式(Nelder,1998)。这类扩展也被称为广义线性混合模型,计算起来可能费时费力,即使如 SAS 中的 Proc NLMIXED 使用了最先进的普适高斯积分法(adaptive Gaussian quadrature)也仍需特别小心(Lesaffre and Spiessens,2001)。在此,我们暂不考虑计算与估计这类复杂的问题,而主要关注模型的比较,我们假定数据分析者可应用先进的计算机程序进行估计并遵

循其程序的规则。在统计分组比较的脉络中,我们既考察固定效应模型,也考察随机效应模型。

多水平分析的灵活性允许数据分析者在建模中一类分组嵌套另一类分组,出来的模型会涉及三个或更多水平。在一个两水平分析中,可将较高水平的分组视为固定的,而将较低水平的分组视为具有随机效应的,研究者还可再把更高水平的分组视为外部变量,从而在更高水平的单位上比较这个两水平模型。这种方法等同于包含了固定效应和随机效应的三水平混合模型,本章中我们也将讨论这种处理方法。

## 10.2　多水平分析导论

### 10.2.1　多水平情景

在社会和行为科学中,嵌套的观察现象已经激发了对层次模型的多水平方法的兴趣,并促使其快速发展。社会学中已经发展出来的情境分析(contextual analysis)便是用来处理使用多水平模型的研究者面临的相同问题。在生物医药科学中已经发展出了被称为“混合效应模型(mixed-effects models)”的方法。在经济学,横截面数据与时间序列数据的结合和分析也与此问题有关。近年来,多水平模型软件的广泛普及已经使得多水平分析更为容易了。

通常,我们收集到的数据具有一种层次结构(hierarchical structure)或聚类结构(clustered structure)。例如,同一个家庭的儿童在生理、智能和行为特征上比从总人口中随机抽取的人更为相似。在同一个班或同一个学校的学生身上,我们也能看到类似情况。本章的第一例就是分析小学的数据,这个例子是关于教育社会学的,但技术和原则可适用于更一般的情形中。

我们将数据的层次视为由在不同水平上分组的单位构成的。这样在一个两水平结构中,儿童或学生可以是第 1 水平的单位,而家庭或班级是第 2 水平的单位:儿童或学生是在第 2 水平单位的家庭或班级中聚类的第 1 水平的单位。层次并不一定止于第 2 水平。我们可以进一步将家庭视为在邻里或社区中聚类的单位,或者班级视为在学校中聚类的单位。邻里或学校则成为层次结构中的第 3 水平。虽然下面我们将主要关注两水平的多水平模型,但是将其一般化便可直接应用于三水平模型。

多水平数据结构并不限于观察情形。设计实验也可以产生数据层次,如临床试验也可以在一些随机选取的医疗中心或人群中进行。这样,数据分析者就可采用多水平模型来解释这种层次结构以便评估出真实的治疗效果。在这些临床试验研究中,被试者是第 1 水平的单位,而它又是嵌套于第 2 水平的单位(即研究中选择的医疗中心)中的。

通常,在这些第 2 水平(或更高水平)的组中(如家庭、选择的学校和国家),其内部成员都具有许多共同的特征。这些单位可以是固定的(如国家),也可以是从随机样本中选取的(如家庭或学校)。但是,在另一些情况下,至少起始分组并不是从具有明显共同特征的个人中产生的,例如,儿童被分配到公立的小学中,再如病人被分配到诊所中。但是,若长时间生活在这些组中,成员也会被塑造出一些前所未有的共有特征。多层次分析对于这两种情形都是适用的。

传统上,研究者对多水平数据的分析要么是在个体水平上进行(忽略了第 2 水平中的变异),或者只在宏观水平上进行(将数据进行了汇总)。但这两种情况都会产生严重的问题。汇总数据的分析会产生所谓的"生态谬误"。忽略个体中的聚集则会违背一些统计模型(如广义线性模型)的基本假设:随机误差项的分布是独立同分布的。因此,这两种方法都不可取。

## 10.2.2 书目提要

在不同的领域中,被称为多水平分析、层次线性模型、混合效应模型的统计方法都是为了处理多水平数据结构中的非独立误差而产生。对这一主题做介绍的著作中至少有六本非常值得注意:

1. Bryk, A. S. and S. W. Raudenbush. 1992, *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

2. Goldstein, H. 1995. *Multilevel Statistical Models*. London: Edward Arnold.

在 http://www.arnoldpublishers.com/support/Goldstein.htm 上可以得到电子版本。

3. Hox, J. J. 1994. *Applied Multilevel Analysis*. Amsterdam: TT-publikaties. 在 http://www.ioe.ac.uk/multilevel/amaboek.pdf 上可以得到电子版本。

4. Kreft, I. G. and J. de Leeuw. 1998. *Introducing Multilevel Modeling*. Thousand Oaks, CA: Sage.

5. Longford, N. T. 1993. *Random Coefficient Models*. New York: Oxford University Press.

6. Snijders, T. and R. Bosker. 1999. *Multilevel Analysis: An introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

如果读者对这一问题的入门性读物感兴趣,第 3 本和第 4 本将是很好的选择;第 2 本和第 5 本包括了相当多的数学背景和细节;第 1 本是对层次线性模型进行了广泛而深入的探讨。清单中的最后一本既提供了详细的入门性知识,也就一些较为高深的问题(如层次对数单位模型)提供了最新的信息。

## 10.2.3 固定效应与随机效应

在后面要讨论的随机截距模型和随机系数模型中,其截距与系数,或者更一般化地称其为参数,都既可处理为固定效应,也可处理为随机效应。并不存

在一种更为合适而快速有效的规则来决定究竟采用固定效应还是随机效应的方法。但是,数据结构的一些特性是应当加以考虑的。

在决定选择固定效应模型还是随机效应模型时,必须考虑的一个因素是第 2 水平单位的数量。斯奈德和博斯科(Snijders and Bosker, 1999)讨论了一种取决于数据中的组数 J 的经验法则。如果 J 较小( $J < 10$ ),则采用固定效应方法;这时如果把“分组”视为从总体中抽取的样本,那么,此时的数据中只包含总体的极少量信息。如果 J 较大( $J \geq 10$ )且每组容量不大( $n_j < 100$ ),则采用随机效应方法,这是因为把 10 组或更多组作为独特或固定的实体处理起来通常数量太多了。如果组的容量较大( $n_j \geq 100$ ),那么,无论采用哪种方法差别都不大。斯奈德和博斯科(Snijders and Bosker, 1999)指出:这一经验法则仅仅是一种第一感觉(a first hunch),而不可作为选择固定效应模型和随机效应模型的决定性因素。

在选择固定效应模型还是随机效应模型上,朗福德(Longford, 1993)推荐应考察第 2 水平单位在“可交换性(exchangeability)”方面的特性。例如,不同的学校和医院是可交换的,因为我们希望得到的推论不依赖于它们的具体身份。依据同一种论证,嵌套于学校和医院之内的第 1 水平单位同样也是可交换的,即学生和病人也是可交换的,这是因为我们希望得到的有关行为或医药方面的推论并不依赖于他们的个体身份。这并不意味着,第 1 水平的单位或第 2 水平的单位在它们中间是完全相同的。在第 1 水平的单位上,其独特性由社会经济地位之类的测量来表示,在第 2 水平的单位上,其独特性由学校或医院类型(如公立还是与教会相关的)来表示。在这类情况下,第 2 水平的单位最好处理为随机的。

另外,国家、种族群体和性别是不可交换的;每个都有其自己独特的性质。每个群体都可以清晰地限定和识别,因此,这类单位最好视为固定的。

可交换性标准和样本量法则常常不谋而合。例如,当我们比较国家、种族群体和性别时,大部分情况下都不会有太多组( $J < 10$ )。在前面章节中,我们所讨论的组间比较都假定第 2 水平单位是固定的。

## 10.3　线性多水平模型的基础知识

### 10.3.1　基本的数据结构

在前一节中,我们已经描述了存在多水平数据结构和需要多水平模型的情形。试考虑一个两水平结构,它描述的是学校之内有聚集在一起的学生,或者邻里之内有聚集在一起的居民。此类两水平数据(单向形式)的结构可用表 10.1 来汇总。

本章所采用的符号遵循多水平分析的文献中的惯例(如 Bryk and

Raudenbush, 1992; Goldstein, 1995)。因此,我们用下标" j"表示第 2 水平单位,但如果为了与前面章节一致, j可以很方便地换成 g 来代表观察属于的组类。注意每组或每个第 2 水平单位不必有相同的第 1 水平观察数 $n_j$。

表 10.1 两水平数据的结构

| 第 2 水平单位 | 第 1 水平单位 | | | |
|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{21}$ | … | $y_{n_1 1}$ |
| 2 | $y_{12}$ | $y_{22}$ | … | $y_{n_2 2}$ |
| 3 | $y_{13}$ | $y_{23}$ | … | $y_{n_3 3}$ |
| … | … | … | | … |
| J | $y_{1J}$ | $y_{2J}$ | … | $y_{n_J J}$ |

注:每一行代表一个第 2 水平单位 j,其中, i代表在第 2 水平单位之内第 1 水平单位的序号,第 1 水平单位由下标" ij"来确定。第 2 水平单位 j中包括了 $n_j$ 个基本观察,第 2 水平单位的总数为 J。

如果第 2 水平单位是嵌套于更高水平单位中时,就会出现一个三水平数据结构。想象表 10. 1 是一个更大表的分表,这个大表有 K 个第 3 水平单位。沿用前面的例子,学校是嵌套于学区范围内的;邻里是嵌套于镇或城市中的。同样,每个分表也不必有相同的行数(第 2 单位数)和列数(第 1 水平单位数)。

## 10.3.2 随机截距模型

### 10.3.2.1 单向 ANOVA

首先我们来考察一个没有解释变量的模型,形式如下:

$$y_{ij} = \beta_{0j} + e_{ij} \tag{10.1}$$

其中, $y_{ij}$ 表示我们关心的一个结果变量, $e_{ij}$ 表示未予解释的误差或残差, i 和 j 分别表示第 1 水平和第 2 水平单位的序号。按照惯例,我们假定 $e_{ij} \sim N(0, \sigma^2)$, $\sigma^2$ 为第 1 水平的方差。这个模型实际上就是包含了随机效应的单向 ANOVA,截距项 $\beta_{0j}$ 就是各组的均值。

在第 2 水平(组的水平)上,依变量在各组的均值 $\beta_{0j}$ 又是总均值 $\gamma_{00}$ 加上随机误差项 $u_{0j}$ 的函数:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{10.2}$$

在此,我们定义 $E(\beta_{0j}) = \gamma_{00}$, $var(\beta_{0j}) = \tau_{00}$,假定 $u_{0j} \sim N(0, \tau_{00})$。我们称 $\tau_{00}$ 为第 2 水平的方差。多水平分析中的一个重要概念"类内相关系数"(intraclass correlation coefficient)与 $\tau_{00}$ 有关,其公式为:

$$\rho = \tau_{00} / (\tau_{00} + \sigma^2) \tag{10.3}$$

所有的多水平线性模型都可分解为两个组成部分——固定部分和随机部分。固定部分由模型中的参数和解释变量代表(虽然本例中没有解释变量),

随机部分包括两个方差 $\sigma^2$ 和 $\tau_{00}$ 。

10.3.2.2　单向 ANCOVA

下面我们将扩展式(10.1)和式(10.2)中的模型,在第 1 水平方程中加入一个解释变量或协变量:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij} \tag{10.4}$$

现在有两个第 2 水平方程,每一个方程对应一个 $\beta$,即:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \end{aligned} \tag{10.5}$$

注意,$x_{ij}$ 的效应被约束为对每个第 2 水平单位都相同。从线性回归的角度来看,这意味着各组的回归线彼此平行。这样,式(10.4)和式(10.5)定义了一个包含随机效应的 ANCOVA 模型。

模型可以继续扩展为包括第 2 水平上的协变量。接下来,我们便考察此类扩展,并区分出组内回归(within-group regression)和组间回归(between-group regression)。在第 2 水平的随机截距方程中加入一个第 2 水平的变量:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + u_{0j} \tag{10.6}$$

在此 $w_{1j}$ 可被视为对各组的结果均值进行的估计。式(10.4)和式(10.5)中设定的其他两个方程保持不变。

经典的 ANOVA 假定协变量效应 $\gamma_{10}$ 在每一组中都相同。使用矩阵的符号,此模型可以很容易地推广为第 1 水平和第 2 水平方程中都有多个解释变量的模型。在此,我们用列向量来表示每个组(即每个第 2 水平单位 j)中的第 1 水平单位,这样便可在第 1 水平方程中省略下标“ i”;对于从第 1 水平方程中得到的每个 $\beta$,我们用列向量来代表第 2 水平单位 j,这样便可在第 2 水平方程中省略下标“ j”。

$$\begin{aligned} y_j &= X_j\beta_j + e_j \\ \beta_k &= W_0\gamma_k + u_0 \end{aligned} \tag{10.7}$$

其中,$\beta_k$ 是由 J 个随机系数组成的向量,每个系数都是一个 $\beta_k$ ;$\gamma_k$ 是对每个第一水平的参数 $\beta_k$ ( $k = 0,\cdots,K$ )进行第二水平的回归时的参数所组成的向量;$u_0$ 是随机误差向量,当 $k = 1,\cdots,K$ 时元素全部为零,当 $k = 0$ 元素服从分布 $(0,\tau_{00})$ ,故用 $u_0$ 代表;$X_j$ 和 $W_0$ 包括了各自的解释变量。当 $k = 1,\cdots,K$ 时,$W_0$ 包含的是标量常数 1;当 $k = 0$ 时,它可以有自由参数进行估计。

到目前为止,我们只考虑了随机截距多水平模型。在下一小节中,我们放宽假定,将讨论包含随机系数的多水平模型。

## 10.3.3　随机系数模型

式(10.7)中的一般随机截距模型假定第 1 水平的回归方程的斜率在各个组中都相同。放宽这一假定将产生下面将要介绍的子模型:

#### 10.3.3.1 不存在跨水平效应的随机系数模型

现在我们允许第 1 水平回归方程中的斜率在各组中不同：

$$\begin{aligned} y_j &= X_j\beta_j + e_j \\ \beta_k &= W_0\gamma_k + u_k \end{aligned} \tag{10.8}$$

其中，$u_k$ 除了当 $k = 1,\cdots,K$ 时不再被限定为零以外，与前面所定义的随机误差向量完全相同。其他各项的定义与前面完全相同。当 $k = 1,\cdots,K$ 时，此模型不包括任何的第 2 水平的协变量，而是让第 1 水平的系数在各个组（即在各个第 2 水平单位中）中有所不同。在这种意义上讲，此模型不含有所谓的跨水平系数，即不含有第 1 水平和第 2 水平协变量间的交互作用。

如果分组被视为固定，则此模型类似于通常加入所有 x 变量与分组变量交互作用的方法。将 X 与分组变量相乘，可得：

$$y_j = X_j\beta + u_{0j} + X_j\gamma_j + e_j \tag{10.9}$$

在许多应用中，研究者只考虑分组的主效应的向量（代替用 $u_{0j}$ 代表的随机部分）和 $\gamma_j$ 的交互效应。因此，此模型可视为随机系数模型的固定效应形式。

#### 10.3.3.2 存在跨水平效应的随机系数模型

现在我们把前一节中的模型进行进一步扩展，即把协变量和随机误差都包括在每个第 2 水平方程中。用矩阵符号表示，可得：

$$\begin{aligned} y_j &= X_j\beta_j + e_j \\ \beta_k &= W_k\gamma_k + u_k \end{aligned} \tag{10.10}$$

此模型中我们不仅允许用随机误差来解释截距和系数，而且还允许用一组 w 变量来解释每个 $\beta_k$ 在各组中的变异。换言之，这些 w 变量与 x 变量一起构成了跨水平效应。事实上，式（10.10）定义了层次线性模型的一般形式，而前面的模型都可视为其中的某些因素在约束条件下的特例。

### 10.3.4 示例

为了加深对采用线性多水平模型来分析多水平数据的理解，让我们来看一个例子。下面我们分析的数据是关于荷兰 11 岁的小学生（Snijders and Bosker，1999），可以从汤姆• 斯奈德的主页上得到（http://stat.gamma.rug.nl/）。有着完整记录的样本量是来自 131 个班级的 2 287 名学生。在排除了缺失值之后的数据中，班级的大小从 4 到 35 不等。嵌套结构是班级之内有小学生。

表 10.2　语言测试成绩数据的一部分

| | 个体水平的数据 | | |
|---|---|---|---|
| 学生 | 智商 | 语言测试成绩 | 社会经济地位 |
| 1 | 3.165 90 | 46 | −4.812 00 |
| 1 | 2.665 90 | 45 | −17.812 00 |
| 1 | −2.334 10 | 33 | −12.812 00 |
| … | … | … | … |
| 2 | −0.334 60 | 21 | −12.812 00 |
| 2 | −1.334 10 | 27 | −12.812 00 |
| 2 | −3.834 10 | 16 | −7.812 00 |
| … | … | … | … |
| 10 | −1.334 10 | 28 | −12.812 00 |
| 10 | 1.334 10 | 33 | −7.812 00 |
| 10 | 0.665 94 | 38 | −7.812 00 |
| … | … | … | … |
| | 班级水平的数据 | | |
| 班级 | 平均智商 | 是否年级混合 | 班级大小 |
| 1 | −1.514 10 | 0 | 5.9 |
| 2 | −2.834 10 | 1 | −14.1 |
| 10 | −1.334 10 | 1 | −13.1 |
| … | … | … | … |

资料来源： http://stat.gamma.rug.nl/。

依变量是在语言测试中的得分。我们感兴趣的问题是，得分是如何有赖于学生的智商、家庭社会经济地位以及在班级水平上的一些变量的。数据包含在两个原始数据文件中，一个是学生水平的（第 1 水平），另一个是班级水平的（第 2 水平）。表 10.2 介绍了两个数据文件的一部分。

完整的数据文件可以从表 10.2 的资料来源 http://stat.gamma.rug.nl/得到，或者在 http://www.staff.uiuc.edu/~f-liao/multilevel/asaeg11.dat 和 http://www.staff.uiuc.edu/~f-liao/multilevel/asaeg12.dat 得到。通常情况下，数据文件是矩形的，第 2 水平的编号被列为第一个变量。下面我们将使用 HLM 5 来估计与 10.3.2 和 10.3.3 节中介绍的模型相对应的一系列多水平模型。

## 10.3.5　存在随机效应的 ANOVA

我们估计的第一个模型为：

$$LANGSC_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

其中，LANGSC 是学生在语言测试中的成绩。如前面所述，多水平模型可以分

解为固定部分和随机部分，如表 10.3 所示。在第一个模型（空模型）中，固定部分只有一个参数 $\gamma_{00}$，估计值为 40.36。随机部分由两个方差部分组成：$\sigma^2 = 64.56$；$\tau_{00} = 19.63$。从这些参数中，我们可得到 ρ 的一个估计值为 $19.63/(19.63+64.56)=0.23$。这个值相当高：一般教育研究得到的 ρ 值范围为 0.05～0.20。对 ρ 存在两种解释：从第 1 水平的角度来看，这一估计显示了班级内学生之间的相关程度。另一种解释是，这一估计测量了结果变量在第 2 水平单位之间的方差所占的比例。

**表 10.3　对存在随机效应的 ANOVA 模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 40.362 | 0.427 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 19.633 | 4.431 |
| $\sigma^2$ | 64.564 | 8.035 |
| 总离差 | 16 253.081 | |

## 10.3.6　存在随机效应的 ANCOVA

现在我们对前面的 ANOVA 模型进行扩展，在第 1 水平的方程中加入一个学生层面的解释变量——语言智商（IQ）：

$$LANGSC_{ij} = \beta_{0j} + \beta_{1j} IQ_{ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

这样我们则有两个第 2 水平的方程，与第 1 水平方程中的每个 β 一一对应。到目前为止，我们考虑的只是随机截距多水平模型。也就是说，我们只允许截距项随机变动，而将解释变量的系数固定为在每组中都相同。对于智商影响 $\gamma_{10}$ 的参数估计为 2.49，标准误为 0.08，估计值高度显著。从表 10.4 中的结果计算出的类内相关系数下降为 0.19。

所有组的斜率相同，而从第一个第 2 水平方程得到的截距是变动的，这表明回归线是彼此平行的，且与 Y 轴相交于不同点。

**表 10.4　对存在随机效应的 ANCOVA 模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 40.608 | 0.305 |
| $\gamma_{10}$ | 2.488 | 0.081 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 9.601 | 3.099 |
| $\sigma^2$ | 42.245 | 6.500 |
| 总离差 | 15 253.934 | |

现在我们进一步扩展模型,在第 2 水平方程中的随机截距方程中加入一个协变量:

$$LANGSC_{ij} = \beta_{0j} + \beta_{1j}IQ_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}GIQ_{j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

其中 ,$GIQ_j$ 是每个班级的平均智商。显而易见,分班级的平均智商可以解释截距项的相当一部分差异,$\gamma_{01}$ 的估计值为 1.59,标准误为 0.32(见表 10.5)。$\rho$ 的估计值进一步降到了 0.16。组内回归系数 $\gamma_{10}$ 为 2.41,组间回归系数为组内回归系数 $\gamma_{10}$ 与组均值回归系数 $\gamma_{01}$ 之和,即 2.41 + 1.59 = 4.00。平均而言,具有某一智商水平的学生在平均智商水平较高的班级里可以在语言测试里获得更高的分数(幅度为 1.59)。这样,班级中对智商的回归斜率仍然彼此平行,但是组间回归斜率变得更陡峭了。在其他研究情境下事实也可能相反,也就是说,组间回归斜率线比组内回归线更为平缓。

**表 10.5 对存在随机效应的扩展 ANCOVA 模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 40.741 | 0.284 |
| $\gamma_{01}$ | 1.589 | 0.320 |
| $\gamma_{10}$ | 2.415 | 0.085 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 7.886 | 2.808 |
| $\sigma^2$ | 42.172 | 6.494 |
| 总离差 | 15 232.190 | |

## 10.3.7 不存在跨水平效应的随机系数模型

下面我们将对前面的模型进行进一步扩展,在对 $\beta_{1j}$ 的第 2 水平方程中包括随机成分:

$$LANGSC_{ij} = \beta_{0j} + \beta_{1j}IQ_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}GIQ_{j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

如果分组被视为固定的,事实上此模型就是前面章节中介绍的混合效应随机系数模型,它等同于第 6 章中的一个包含分组变量与 x 变量的交互作用的广义线性模型估计。对模型的估计(见表 10.6)给出了智商对语言测试成绩的影响($\gamma_{10}$)为 2.46,这一斜率的标准差(即 $\sqrt{var(u_{1j})}$)为 0.46。平均效应的取值范围(±2 个标准差)为 1.54 ~ 3.38。这表明,虽然智商总是对语言成绩有正向影响,但是较大的影响是较小影响的 2 倍多。在本例中,组间回归线要比一些组内回归线更为陡峭,而比另一些平缓,这是由于现在的组内回归

斜率彼此不同。

**表 10.6 对不存在跨水平效应的随机系数模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 40.750 | 0.286 |
| $\gamma_{01}$ | 1.412 | 0.345 |
| $\gamma_{10}$ | 2.460 | 0.086 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 8.077 | 2.842 |
| $\tau_{11}$ | 0.212 | 0.460 |
| $\sigma^2$ | 41.339 | 6.430 |
| 总离差 | 15 217.874 | |

## 10.3.8 存在跨水平效应的随机系数模型

我们进一步扩展模型，在两个第 2 水平方程中加入一个第 2 水平变量——班级大小(GSIZE)：

$$LANGSC_{ij} = \beta_{0j} + \beta_{1j}IQ_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}GIQ_{j} + \gamma_{02}GSIZE_{j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}GSIZE_{j} + u_{1j}$$

参数 $\gamma_{11}$ 描述的就是班级大小与个体智商之间的跨水平交互作用。班级大小的范围为 5 ~ 37。负的估计值表明在小班中智商对语言测试成绩的影响更大(见表 10.7)。

**表 10.7 对存在跨水平效应的随机系数模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 40.923 | 0.289 |
| $\gamma_{01}$ | 1.224 | 0.350 |
| $\gamma_{02}$ | 0.072 | 0.039 |
| $\gamma_{10}$ | 2.444 | 0.084 |
| $\gamma_{11}$ | -0.025 | 0.012 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 7.919 | 2.814 |
| $\tau_{11}$ | 0.199 | 0.446 |
| $\sigma^2$ | 42.310 | 6.427 |
| 总离差 | 15 222.972 | |

上面的模型是线性多水平模型的一般形式的示例，只要模型可以识别，研究者则可以加进尽可能多的变量。在本例中，我们可以加入数据中的所有变量。但是这样做在统计上无效，在实质理论上也无意义。一种通行的办法是先加入所有合理的变量，然后再使用一种准则(如在广义线性模型中的 LRT

准则)来简化模型,使其更为简洁。下面我们将介绍一个如此简化过的模型:

$$LANGSC_{ij} = \beta_{0j} + \beta_{1j}IQ_{ij} + \beta_{2j}SES_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}GIQ_j + \gamma_{02}MGRADE_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}MGRADE_j + u_{1j}$$
$$\beta_{2j} = \gamma_{20}$$

模型中加入了一个新的第 1 水平变量“SES”(社会经济地位)和一个新的第 2 水平变量“MGRADE”(表示班级中是否有混合年级的状态),而忽略了第 2 水平变量 GSIZE。

同前面的例子一样,我们使用完全最大似然估计(以便进行 LRT 检验)而非限制似然估计,模型得到的偏差(deviance)[①]的分值为 15 086. 33(见表 10. 8)。在所有三个第 2 水平方程中加入 GIQ,GSIZE,MGRADE 以及随机误差的模型通过完全似然估计得到的总离差分值为 15 079. 58。简化模型对数据的拟合度降低得很小,LRT 检验值为 15 086. 33 - 15 079. 58 = 6. 75,自由度为 19 - 10 = 9。我们节省了 9 个自由度,但在解释结果的变异程度上只损失了一点。包含混合年级变量的估计值表明:在混合年级的班级中,学生普通在语言测试中成绩较差,而在这些混合年级的班级中智商的影响更大些。因此,在班级的混合年级状态与个体智商之间存在显著的跨水平效应。

**表 10. 8 对经过简化的存在跨水平效应的随机系数模型的 ML 估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 41. 453 | 0. 330 |
| $\gamma_{01}$ | 0. 833 | 0. 328 |
| $\gamma_{02}$ | -1. 714 | 0. 614 |
| $\gamma_{10}$ | 2. 079 | 0. 090 |
| $\gamma_{11}$ | 0. 542 | 0. 179 |
| $\gamma_{20}$ | 0. 156 | 0. 014 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 7. 518 | 2. 742 |
| $\tau_{11}$ | 0. 118 | 0. 343 |
| $\sigma^2$ | 39. 281 | 6. 267 |
| 总离差 | 15 086. 328 | |

## 10. 3. 9 线性多水平模型的假设

线性多水平模型的基本假设是对线性回归模型假设的扩展。它们包括:

**设定**:模型中的固定部分和随机部分是否已经包括了所有必要的变量?

**误差分布**:在所有水平上,随机误差项是否都服从正态分布(包括在更高

① 即 LRT 统计量。——译者注

水平单位中的第 1 水平的随机误差分布)?

**方差同质性**:在所有水平上,从协变量角度看,随机误差是否有恒定的协方差?

**变量和残差的独立性**:在分析的每一水平上,变量和残差是否彼此不相关?

**跨水平独立性**:某一水平上的随机误差是否独立于其他水平上的随机误差? 更进一步,它们是否也独立于其他水平上的变量?

要满足线性多水平模型的基本假设,就要对上述问题都做出肯定回答。

## 10.4 广义线性多水平模型的基础知识

在实际应用中,通常我们会遇到离散的依变量,如二分变量、定类变量、定序变量或者事件发生数目。我们运用在第 6 章和第 7 章中所讨论的广义线性模型(GLM),前面各节中讨论的模型可以扩展为层次广义线性模型或者广义线性多水平模型(generalized linear multilevel models,以下简称 GLMM),由此便可以处理此类非连续的依变量。对于 GLMM 的估计问题已得到解决并经过详细讨论(如 McCullogh, 1997; Schall, 1991)。许多多水平软件也可对 GLMM 进行估计。GLMM 还可以扩展为包括潜变量,这便形成了被称为广义线性潜混合模型(generalized linear latent and mixed models,以下简称 GLLAMM)。此类模型可以处理包括连续、二分、定序或非定序的类别变量、事件数目、生存数据及序列数据等多种类型的多变量响应。

沿用我们前面章节中的符号,GLMM 在两水平情境下的第 1 水平方程可写为:

$$y_j = \mu_j + e_j \tag{10.11}$$

其中, j是第 2 水平单位的序号,其他符号与第 6 章中的定义相同。如前所述,令 $g(\cdot)$ 为单调连续函数,满足以下条件:

$$g(\mu_j) = \eta_j = X_j \beta_j \tag{10.12}$$

但是,对于向量 $\beta$,我们需要进一步定义:

$$\beta_j = W_j \gamma_j + u_j \tag{10.13}$$

举例来讲,如果第 1 水平数据是二分的,联系函数为典范对数单位,对于两水平数据结构我们可得:

$$y_j = P_j + e_j \tag{10.14}$$

$$\mathrm{In} \frac{P_j}{1 - P_j} = X_j \beta_j \tag{10.15}$$

$$\beta_j = W_j \gamma_j + u_j \tag{10.16}$$

其中, $P_j = P(y_j = 1)$ 。其他的第 1 水平数据分布和相应的连接函数也可以很方便地写出。

### 10.4.1　存在跨水平效应的随机系数对数单位模型

我们仍用荷兰学校的例子，但是把语言测试成绩变量转化为一个虚拟变量，如果学生成绩在 30 分及以上，编码为 1。

表 10.9 中拟合的模型是对表 10.8 中模型的简化：

$$\ln\left[\frac{P(LANGPASS_{ij}=1)}{1-P(LANGPASS_{ij}=1)}\right]=\beta_{0j}+\beta_{1j}IQ_{ij}$$

$$\beta_{0j}=\gamma_{00}+\gamma_{01}MGRADE_j+u_{0j}$$

$$\beta_{1j}=\gamma_{10}+\gamma_{11}MGRADE_j+u_{1j}$$

结果是用 HLM 5 得到的，使用的是从总体平均模型（population average model）中得到的具有稳定标准误的结果。我们用 SAS PROC NLMIXED 估计了同一模型。HLM 进行的是限制似然估计和完全似然估计，而 SAS 程序则采用以似然为基础的普适高斯积分法进行估计，给出的估计则更为准确，但是估计过程较费时。SAS 给出的估计值如表 10.10 所示。两组所得的结果大致接近，但并不完全相同。

**表 10.9　HLM 对存在跨水平效应的随机系数对数单位模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 2.518 | 0.097 |
| $\gamma_{01}$ | 0.428 | 0.125 |
| $\gamma_{10}$ | 0.551 | 0.040 |
| $\gamma_{11}$ | 0.145 | 0.082 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 0.459 | 0.677 |
| $\tau_{11}$ | 0.007 | 0.085 |
| 总离差 | 996.957 | |

**表 10.10　SAS 对存在跨水平效应的随机系数对数单位模型的估计**

| 固定效应 | 估计值 | 标准误 |
|---|---|---|
| $\gamma_{00}$ | 2.775 | 0.142 |
| $\gamma_{01}$ | 0.494 | 0.134 |
| $\gamma_{10}$ | 0.588 | 0.057 |
| $\gamma_{11}$ | 0.185 | 0.098 |
| 随机效应 | 方差的各组成部分 | 标准差 |
| $\tau_{00}$ | 0.544 | 0.232 |
| $\tau_{11}$ | 0.013 | 0.027 |
| $\tau_{01}$ | −0.047 | 0.063 |
| −2£ | 1 274.5 | |

接下来,我们研究一个国家被作为第 2 水平单位的多水平示例。克卢格尔、梅森、韦格纳(Kluegel, Mason, and Wegener, 1995)就社会公正问题对 11 个欧洲国家以及日本和美国进行了调查。他们提出的问题之一是关于政府在个人收入限制中的作用。被访者被要求对下面陈述给出意见:“政府应该对每个人能够赚到的钱数设置一个上限”。原本的五类被压缩成了两类:1 代表同意,0 代表不同意,因为总体而论存在着强烈的两极化分布趋势。对于第 1 水平模型,我们只使用了两个解释因素——性别和教育(用三个虚拟变量来代表)。第 2 水平模型中的两个方程都只包含了一个解释变量——1991 年的失业率:

$$\ln\left[\frac{P(GVTLIMIT_{ij}=1}{1-P(GVTLIMIT_{ij}=1}\right]=\beta_{0j}+\beta_{1j}GENDER_{ij}+\beta_{2j}MEDVOC_{ij}$$

$$+\beta_{3j}2NDEDUC_{ij}+\beta_{4j}3RDEDUC_{ij}$$

$$\beta_{01}=\gamma_{00}+\gamma_{01}UNEMPLOY_j+u_{0j}$$

$$\beta_{1j}=\gamma_{10}+\gamma_{11}UNEMPLOY_j+u_{1j}$$

其中,GVTLIMIT 是二分依变量;若为男性则 GENDER 取值为 1,若为女性则 GENDER 取值为 0;三个教育变量分别表示被访者接受了初等职业教育、中等教育、高等教育,而以小学及以下教育程度为参照类别。UNEMPLOY 表示失业率,是唯一的第 2 水平变量。个案数据共有 16 378 个,本次分析和下面的分析都将基于这些数据。用 SAS PROC NLMIXED 估计出的结果显示在表10.11中,而估计有对数单位联系函数的 GLMM 的 SAS 语句在 10. B. 1 节中给出。

表 10.11 SAS 对存在跨水平效应的两个 GLMM 的估计

| | Logit | | Probit | |
|---|---|---|---|---|
| 固定效应 | 估计值 | 标准误 | 估计值 | 标准误 |
| $\beta_{2j}$ | 0.551 | 0.045 | 0.340 | 0.027 8 |
| $\beta_{3j}$ | 0.864 | 0.050 | 0.532 | 0.030 |
| $\beta_{4j}$ | 0.990 | 0.049 | 0.609 | 0.030 |
| $\gamma_{00}$ | -1.425 | 0.319 | -0.872 | 0.195 |
| $\gamma_{01}$ | 5.113 | 2.212 | 3.105 | 1.351 |
| $\gamma_{10}$ | 0.724 | 0.137 | 0.446 | 0.083 |
| $\gamma_{11}$ | -1.743 | 0.953 | -1.085 | 0.575 |
| 随机效应 | 方差的各组成部分 | 标准误 | 方差的各组成部分 | 标准误 |
| $\tau_{00}$ | 0.277 | 0.112 | 0.103 | 0.042 |
| $\tau_{11}$ | 0.036 | 0.021 | 0.013 | 0.008 |
| $\tau_{01}$ | -0.032 | 0.036 | -0.012 | 0.013 |
| -2£ | 20 766 | | 20 764 | |

## 10.4.2 存在跨水平效应的随机系数概率单位模型

我们也可以拟合一个联系函数为概率单位的 GLMM。

$$\Phi^{-1}(P(GVTLIMIT_{ij}=1))=\beta_{0j}+\beta_{1j}GENDER_{ij}+\beta_{2j}MEDVOC_{ij}$$

$$+ \beta_{3j}2NDEDUC_{ij} + \beta_{4j}3RDEDUC_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}UNEMPLOY_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}UNEMPLOY_j + u_{1j}$$

在表 10.11 中,概率单位模型的估计结果与对数单位模型的估计结果显示在相邻的列中。对于固定效应参数来说,可以非常容易地将概率单位估计转换为对数单位的对应值,只要乘以一个位于 1.6 和 $\sqrt{3}$ 之间的常数,但是对于随机效应方差来说,则不能进行此种转换。

## 10.5　多水平模型中作为外部变量的“组”

在分析中,我们可以再包含另一个分组变量,以形成更高一级的“水平”。例如,在欧洲、美国和日本进行的有关社会正义的跨国研究中就可以包括一个第 3 水平变量,序号为 k,表示它以前是否是共产主义国家。因此在这一水平上就有两个单位,这样进行包含三个水平的分析就显得既无意义也无法操作。但是,我们可以将这个变量处理为外部变量,如同前面章节中一样进行分组比较。

具体来讲,基于 10.4.1 节研究中的所有国家的个体可以建立一个模型,同时可以对两组国家单独构建两个模型,这些模型的对数似然就可用来进行似然比检验,结果为 4.15。在给定水平 α 下的显著性检验表明:这两种过去的意识形态集团之间的差异已经大到应引起单独注意的程度。

这一检验是一般性的,但并非所有参数都必须同时检验。例如,可以把随机部分约束为在各组中有着某一相同的固定值,然后去检验各组的固定效应是否相同,反过来也同样可以。也可同第 8 章和第 9 章一样,只检验某些固定参数(或随机参数)在各组中是否相同。最后,只要分组可以被合理地认为是固定的,而且分组总数不太多又便于处理,外部变量的方法可以在不止一个分析水平上容纳分组。

## 10.6　多水平模型与分组比较之间的关系

### 10.6.1　联结固定效应模型和随机效应模型

让我们以一个简单的两水平模型开始,模型中有一个第 1 水平的解释变量和一个第 2 水平的解释变量:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + u_{1j}$$

如前所述，$E(\beta_{0j}) = \gamma_{00}$，$var(\beta_{0j}) = \tau_{00}$，$E(\beta_{1j}) = \gamma_{10}$，$var(\beta_{1j}) = \tau_{11}$，并且 $cov(\beta_{0j}, \beta_{1j}) = \tau_{01}$。用第2水平的项来替代第1水平方程中的参数，我们可得到一个组合模型：

$$y_{ij} = \gamma_{00} + \gamma_{01} w_j + \gamma_{10} x_{ij} + \gamma_{11} w_j x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{ij} \tag{10.17}$$

其中，$u_{0j}$ 给出了随机截距项，而随机误差 $u_{1j}$ 是 $x_{1j}$ 的随机系数项。

假定第2水平变量 $w_j$ 代表的就是第2水平的分组。由于现在 $w_j$ 是类别型的，我们必须根据不同类别而改变系数，这样 $u_{0j}$ 和 $u_{1j}$ 就被吸纳到了系数 $\gamma_{01}$ 和 $\gamma_{11}$ 中去了，如果我们再给系数加上下标"j"，则新的组合模型形式为：

$$y_{ij} = \gamma_{00} + \gamma_{01j} w_j + \gamma_{10} x_{ij} + \gamma_{11j} w_j x_{ij} + e_{ij} \tag{10.18}$$

其中，j取值为1（第1组）到J（最后一组）。式（10.18）给出了一个固定效应模型，相当于包括了交互作用的单模型的组间差异检验，而这已在第3章和第6章介绍过了。参数 $\gamma_{01j}$ 联合起来检验J组中的截距或均值差异。参数 $\gamma_{11j}$ 联合起来检验J组中的斜率或系数差异。

如果第2水平上的组数非常多，而且不存在单一的特征可以区分各组时，荷兰学校数据的例子中正是如此，这时随机效应模型明显更好一些。在组数较少（如二到五组，如前一节中那样）的研究情境下进行分组比较时，应明智地选择固定效应模型。但是，有时研究问题位于这两种理想情境之间。上面考察过的对于社会正义的跨国比较就是这样一个例子。虽然对于国家可以进行固定效应模型分析，但是研究者并没有包括研究总体中的每个欧洲国家或每个后共产主义国家。在此意义上，是进行了某种（便利）抽样。第2水平单位的数量（13个）也使得固定效应模型有些笨拙。从第1水平单位的数量（每组中都大于100）和第2水平单位的数量（大于10）来判断，选择固定效应模型还是随机效应模型差别很小（Snijders and Bosker，1999）。在下一节中我们将比较对于社会正义数据进行的固定效应模型和随机效应模型分析。

## 10.6.2 示例

首先我们在固定效应假定下对跨国社会正义数据拟合一个对数单位联系的GLMM。模型服从式（10.18）中的合并模型形式，其表达式为：

$$\begin{aligned}
&\ln\left[\frac{P(GVTLIMIT_{ij}=1)}{1-P(GVTLIMIT_{ij}=1)}\right] \\
&= \gamma_{00} + \gamma_{01j}NATION_j + \gamma_{10}GENDER_{ij} + \gamma_{20}MEDVOC_{ij} \\
&\quad + \gamma_{30}2NDEDUC_{ij} + \gamma_{40}3RDEDUC_{ij} + \gamma_{11j}NATION_jGENDER_{ij} \\
&\quad + \gamma_{21j}NATION_jMEDVOC_{ij} + \gamma_{31j}NATION_j2NDEDUC_{ij} \\
&\quad + \gamma_{41j}NATION_j3RDEDUC_{ij}
\end{aligned}$$

其中，类别变量NATION以离差来编码，这样主效应 $\gamma_{10}$ 到 $\gamma_{40}$ 给出的就是系数的平均值。模型是用SPSS估计的。此外，我们用SAS PROC NLMIXED对同一数据估计了随机效应假定下的对数单位连接GLMM。模型如式（10.17）所

示，只是其中不存在任何 w 变量，其表达式为：

$$\ln\left[\frac{P(GVTLIMIT_{ij}=1)}{1-P(GVTLIMIT_{ij}=1}\right]=\gamma_{00}+\gamma_{10}GENDER_{ij}+\gamma_{20}MEDVOC_{ij}$$
$$+\gamma_{30}2NDEDUC_{ij}+\gamma_{40}3RDEDUC_{ij}$$
$$+u_{0j}+u_{1j}GENDER_{ij}+u_{2j}MEDVOC_{ij}$$
$$+u_{3j}2NDEDUC_{ij}+u_{4j}3RDEDUC_{ij}$$

其中，$u_{1j}$ 到 $u_{4j}$ 是第 1 水平系数中的随机效应。为了简化模型和估计，设定 $u_0$，$u_1$，$u_2$，$u_3$，$u_4$ 的协方差为零。10.B.3 给出了相应的 SAS 命令语句。如表 10.12 所示为这两个模型的结果。由此可知，从固定效应模型和随机效应模型得到的参数估计几乎完全相同，只有 $\gamma_{20}$ 的估计值在两个模型间有着显著差异。但是，固定效应模型全部低估了估计值的标准误，不论是“平均”效应还是第 2 水平的效应。在这种情况下，两个模型都是无法用“正确”或“不正确”来评判的。显而易见，选择何种模型取决于研究者本身，如果他强调每个国家的特征具有独特性，就会选择固定效应模型；如果他认为国家是从一个更大的总体中抽取出来的，就会采用随机效应模型。

这个例子还说明了本章开始已讲述的很重要的一点：分组统计比较既可作为单水平的 GLM，也可作为 GLMM 处理。一个单水平的包含分组比较系数的 GLM 等同于包含固定效应的 GLMM。但是，在多水平的模型框架中，效应也可以设定是随机的。

**表 10.12 SAS 对两个存在跨水平效应的 GLMM 模型的估计**

| | 固定效应模型 | | 随机效应模型 | |
|---|---|---|---|---|
| 参数 | 估计值 | 标准误 | 估计值 | 标准误 |
| $\gamma_{00}$ | −0.754 | 0.040 | −0.765 | 0.179 |
| $\gamma_{10}$ | 0.523 | 0.036 | 0.510 | 0.067 |
| $\gamma_{20}$ | 0.465 | 0.053 | 0.516 | 0.069 |
| $\gamma_{30}$ | 0.852 | 0.058 | 0.856 | 0.082 |
| $\gamma_{40}$ | 1.032 | 0.056 | 1.034 | 0.105 |
| 第 2 水平效应 | 沃德 $\chi^2$ | 自由度 | 方差的各组成部分 | 标准误 |
| $\tau_{00}$ | 300.030 | 12 | 0.398 | 0.162 |
| $\tau_{11}$ | 48.224 | 12 | 0.043 | 0.024 |
| $\tau_{22}$ | 34.746 | 12 | 0.029 | 0.025 |
| $\tau_{33}$ | 45.001 | 12 | 0.049 | 0.032 |
| $\tau_{44}$ | 75.335 | 12 | 0.108 | 0.056 |
| −2£ | 20 542 | | 20 742 | |

注：对于固定效应模型，表中报告的是对截距和第 1 水平变量进行联合显著性检验而得到的沃尔德卡方统计量。

## 10.7 多重成员身份模型

有时第 1 水平的单位不能整齐地划分到第 2 水平的单位中去,此情况并不常见。例如,一个学生可能依次上过数个学校(甚至可能同时在数个学校上课),一个人可能属于不止一个家庭。要处理这种复杂情况,就要使用多重成员身份模型(multiple membership models)。这一模型的原理类似于应用模糊集理论(Fuzzy set theory)的模型(Liao, 1989c; Manton, Liu, and Cornelius, 1985; Manton et al, 1987)。也就是说,一个成员只是部分地属于某一集合。

以学校为例,假定我们知道每个学生在每个学校中的成员身份状况。在实际操作中,每个学生 i 的每个成员身份都被赋予了一个权重 $w_{ij}$,对于每个学生来说 $\sum_{j=1}^{J} w_{ij} = 1$。权重可以按照这个学生在每个学校度过的时间长短来给予。如果所有第 1 水平的单位的权重都为 1 时,我们得到的模型就是通常的多水平模型。更为一般地,式(10.7)的第一行被扩展为:

$$Y_j = X_j \beta_j + w_j v_j + e_j \tag{10.19}$$

其中,$w_j$ 是长度 I 的权重向量,$v_j$ 是长度 I 的随机误差向量。更多的细节请参阅相关著作(Hill and Goldstein, 1998; Goldstein, 2000)。

多重会员身份模型并未把个人的成员身份函数定义为非此即彼的(即不是一个 0-1 函数),因此当个人拥有多个身份时,这一模型就有相当大的应用潜力。事实上,这种思维方法可以应用到其他类型的模型比较中,而不只限于多水平模型框架中。

## 10.8 总　结

本章中我们介绍了多水平模型,它在分析中把"分组"处理为一个水平。GLMM 的一般形式为:

$$\begin{aligned} y_{jk} &= g^{-1}(\eta_{jk}) + e_{jk} \\ \eta_{jk} &= X_{jk}\beta_{jk} \\ \beta_{jk} &= W_{jk}\gamma_{jk} + u_{jk} \end{aligned} \tag{10.20}$$

其中,k 表示一个更高水平的分组变量,它可作为第 3 水平处理,也可作为外部分组变量可简便地加以分析。

在利用 GLMM 进行分组比较时,数据分析者必须考虑两个重要的问题:必须决定是否把分组变量处理作为分析中的一个单独水平,必须决定是采用固定效应还是随机效应处理各组之间的差异。这里并不存在正确而快捷的准则,但是单位或分组的可交换性、分析中第 2 水平单位的数量和组的数量、统

计模型的简约性可以作为一般性的指导原则。

# 10. A 用于多水平分析的软件

## 10. A. 1 专用软件

### 10. A. 1. 1 HLM

HLM 的 Windows 版本是最为常见的多水平分析软件,它易于使用,非常便捷。在 HLM 5 版本中,可以进行两水平或三水平的分析,依变量可以是连续的也可以是类别的(二分的、定序的、多项的或者泊松的)。生产 HLM 的科学软件国际公司,其网址为:http://www. ssicentral. com/hlm/hlm. htm,学生版可以在此免费下载。

### 10. A. 1. 2 MIXREG, MIXOR, MIXNO, MIXPREG, MIXGSUR

这些程序由在芝加哥的伊利诺大学的唐• 赫德克尔(Don Hedecker)编撰而成,只接受 ASCII 码的数据输入。MIXREG 进行的是存在自相关误差的混合效应线性回归; MIXOR 进行二分或定序结果的多水平分析;MIXNO 估计多水平情境下的多项逻辑斯蒂回归;MIXPREG 估计多水平泊松回归模型;MIXGSUR 分析多水平的时间分组生存数据。在下面网址上可以得到这些软件及更多的相关信息:http://www. uic. edu/ ~ hedeker/mix. html。

### 10. A. 1. 3 MLn / MLwiN

此软件由伦敦教育研究所的研究人员开发,它是最为全面的多水平软件。不同于 HLM,MLn(以及它的 Windows 版本 MLwiN)除了多水平分析之外,还可以进行数据管理、画图以及一些其他的简单统计分析。它可以处理类别依变量,可以进行宏命令操作,也可以分析任意数目的水平。其网址在伦敦大学教育研究所上:http://www. ioe. ac. uk/mlwin/。

### 10. A. 1. 4 VARCL

VARCL 由尼古拉斯• 郎福德(Nicholas Longford)编写,包含 VARCL3 和 VARCL9 两个程序。后者可以分析最多九个水平但只包含固定系数的模型,而前者可以分析包含随机系数的三个水平的模型。依变量可以是正态分布或非正态分布(如泊松分布、二项分布或伽玛分布)。如果想了解更多信息,请访问 http://www. assess. com/VARCL. html。

## 10. A. 2　通用软件

### 10. A. 2. 1　SAS / Proc Mixed

不同于其他通用软件，SAS 可以拟合相当复杂的多水平模型，同时给出相应的统计量。朱迪思·辛格的一篇文章非常好地介绍了 Proc Mixed 的应用方法（Judith Singer，1998，Journal of Educational and Behavioral Statistics 24：323-355）。超出线性模型的部分，如结果变量是离散的模型，则显得更为复杂；要进行此类分析，用户必须运行 SAS 的宏命令 GLIMMIX 或 NLINMIX。宏命令重复运行 Proc Mixed 实现迭代，用限制似然或残差伪似然（residual pseudo likelihood）来进行估计。一种新的 SAS 程序 Proc NLMixed 也可以使用普适高斯积分法来估计 GLMM。该程序更灵活，可以分析固定部分和随机部分都是非线性的情况。这样它便可以估计许多标准的或用户自己定义的非线性混合效应模型。这一程序通过直接将近似似然最大化来估计模型。可参见网址：http://www.sas.com。

### 10. A. 2. 2　BMDP

程序中包括了几个用于多水平分析的模块：BMDP-3V 是用于分析随机截距模型，BMDP-4V 和 BMDP-5V 都是用于分析纵贯性分析的，但后者可以分析随机系数模型。可参见网址：http://www.spssscience.com/Bmdp/。

### 10. A. 2. 3　S-plus / NLE 和 NLME

NLE 和 NLME 分别包括了一系列用于线性和非线性混合效应模型的 S 函数（或 S-plus 函数）。NLE 和 NLME 软件都可适用于 Unix 和 Windows 平台，也都包括在了最新发布的 S-plus 2000 中。但是它不能处理有着类别响应变量的模型。NLME 同时使用最大似然和限制最大似然估计。数据可以依据一个或多个嵌套因素分组。如果想了解更多信息，请参见 http://www.splus.mathsoft.com。

如果你的系统中还没有 NLME，你可以从贝尔实验室单独得到 NLME 软件：http://cm.bell-labs.com。在这个网址上你可以得到 Unix 系统下或 Windows 系统下的 NLME。

### 10. A. 2. 4　Stata

Stata 包含几个用于多水平模型的模块，例如，loneway 用于空模型，xt 用于纵贯性数据，xtreg 用于线性随机截距模型，xtpois，xtnbreg，xtlogit，xtprobit 分别用于多水平的泊松回归、负二项回归、对数单位回归和概率单位回归模型。最后两个模块是基于广义估计方程方法（generalized estimating equations）的。请参见 http://www.stata.com/。

此外，在索菲娅·拉比·赫斯基思（Sophia Rabe-Hesketh）的网址上可以得到一个用于拟合 GLLAMM 模型的 Stata 程序：http://www.iop.kcl.ac.uk/Iop/Departments/BioComp/programs/gllamm.html。

## 10.A.3　用于其他用途的专用软件

### 10.A.3.1　LISREL

LISREL 的最新版本 LISREL 8 可以进行某些类型的多水平分析，但是你必须具有结构方程的思维方式。更多信息请参见网址 http://www.ssicentral.com/lisrel/mainlis.htm。

### 10.A.3.2　Mplus

Mplus 是另一个用于协方差结构分析的程序，它可以用层次模型或者通径分析模型来分析两水平数据。更多信息请参见网址 http://www.statmodel.com/。

### 10.A.3.3　BUGS

这个软件是由剑桥大学和帝国学院的研究者开发的，它可以进行 Gibbs 抽样算法。它运行多水平分析时要求平衡数据（balanced data）[①]。但是它可以非常灵活地估计许多统计模型。更多信息请参见网址 http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml。

# 10.B　用于 GLMM 的 SAS 程序示例

## 10.B.1　产生表 10.11 中的对数单位结果的命令语句

```
options ls = 80 ps = 60; data one; infile 'isjall.dat';
  input country gender bigvtlim jp average unemrate westeur
  postcom centeur easteur medvoc seceduc thdeduc;

proc nlmixed;
  parms gamma00 = 2.5 gamma01 = 0.5 gamma10 = 0.5 gamma11 = 0.1
    su02 = 1 su12 = 1 cu12 = 0.5 beta3 = 0.5 beta4 = 0.5;
  beta0 = gamma00 + gamma01 * unemrate + u0;
  beta1 = gamma10 + gamma11 * unemrate + u1;
```

① 平衡数据即各组中样本量相同的数据。——译者注

```
  eta = beta0 + beta1 * gender + beta2 * medvoc + beta3 * seceduc + beta4 *
thdeduc;
  expeta = exp (eta);
  p = expeta / (1 + expeta);
  model bigvtlim ~ binary (p);
  random u0 u1 ~ norma1 ( [0,0] , [su02, cu12, su12] )subject = country;
run ;
```

## 10. B. 2　产生表 10. 11 中的概率单位结果的命令语句

```
options ls = 80 ps = 60; data one; infile  'isja11.dat';
  input country gender bigvtlim jp average unemrate westeur
  postcom centeur easteur medvoc seceduc thdeduc;

proc nlmixed;
  parms gamma 00 = 2.5 gamma 01 =0.5 gamma 10 =0.5 gamma11 =0.1
    su 02  = 1 su12 =1 cu12 =0.5 beta3  = 0.5 beta4  = 0.5;
  beta 0 =gamma 00 +gamma 01^n emrate  + u 0;
  beta1 = gamma10  + gamma11^n emrate + u1;
  eta  = beta0 +beta1  *  gender +beta2  *medvoc +beta3  *  seceduc +beta4 *
  thdeduc;
  p  = probnorm (eta);
  model bigvtlim ~binary (p);
  random u 0 u1 ~  norma1( [0,0] , [ su 02, cu12, su12] )  subject  =country;
run;
```

## 10. B. 3　产生表 10. 12 中的包含随机效应的 GLMM 的命令语句

```
options ls  = 80 ps  = 60; data one; infile  'isjall.dat';
  input country gender bigvtlim jp average unemrate westeur
  postcom centeur easteur medvoc seceduc thdeduc;


proc nlmixed;
  parms gamma 00  = 2.5 gamma10  =0.5 gamma 20  = 0.5 gamma 30  =0.5gamma 40  =
0.5
    su 02  = 1 su12  = 0.5 su 22  =0.5 su 32 =0.5 su 42 =0.5;
  beta0  = gamma 00 +u 0;
  beta1 = gamma 10 +u 1;
  beta2  = gamma 20  + u 2;
  beta3  = gamma 30, +u 3;
```

```
  beta4 = gamma40 + u4;
  eta =beta0 + beta1 * gender + beta2 * medvoc + beta3 * seceduc + beta4 *
thdeduc;
  expeta = exp(eta);
  p = expeta / (1 + expeta);
  model bigvtlim ~ binary(p);
  random u0 u1 u2 u3 u4
    ~ normal([0,0,0,0,0], [su02,0, su12,0,0, su22,0,0,0,
    su32,0,0,0,0, su42])
    subject = country;
run;
```

# 术语英汉对照表

Analysis of covariance (ANCOVA), 协方差分析
Analysis of variance (ANOVA), 方差分析
Association, 关联
Assumption, 假设
Asymptotic equivalent, 渐进等效
Asymptotic test, 渐进检验
Asymptotically distribution free (ADF) method, 渐近分布自由方法
Asymptotically normally distributed, 渐进正态分布

Bartlett's test, 巴利特检验
Baseline hazard function, 基准风险函数
Bayes factor, 贝叶斯因子
Bayes theorem, 贝叶斯定理
generalized linear model, 广义线性模型
Bias, 偏误
Binary outcome, 二元结果
Binomial distribution, 二项分布
Bonferroni inequality, 邦弗朗尼不等律
Bootstrap distribution, 自助法分布
Bootstrapping method, 自助法

Canonical link, 典范联系
Canonical parameter, 典型参数
Case-control design (study), 个案—对照设计(研究)
Categorical latent variable, 类别潜变量
Censoring, 截删
Chi-square distribution, 卡方分布
Chi-square statistic, 卡方统计量
$\chi^2$-test, 卡方检验
Chow test, 邹氏检验
Clinical trials, 临床试验
Clustered data, 集群数据
Cochran's test, 科克伦检验
Complementary log-log link, 互补双对数联系
Component effect, 成分效应
Composition factor, 构成因素
Composition structure, 构成结构
Concomitant variable, 伴随变量
Conditional likelihood, 条件似然
Conditional logistic regression, 条件逻辑斯蒂回归
Confidence interval, 置信区间
Confidence level, 置信水平
Confounding factor, 混杂因素
Confounding variable, 混杂变量
Cohort study, 队列研究
Constraint, 约束
Contextual analysis, 情境分析
Contingency table, 列联表
Control group, 对照组
Convergence, 收敛
Correlation matrix, 相关矩阵

Item response theory，项目反应理论
Iterative reweighted least squares（IRLS），迭代重加权最小二乘

Jackknife，刀切法
Joint distribution，联合分布

Kolmogorov-Smirnov test，柯尔莫哥洛夫-斯米诺夫检验

Lagrange multiplier statistic，拉格朗日乘数统计量
Lagrange multiplier test（LMT），拉格朗日乘数检验
Laplace method，拉普拉斯方法
Latent class model（analysis）（LCM），潜类模型(分析)
Latent distribution，潜分布
Latent profile model（LPM），潜层序模型（LPM）
Latent structure model，潜结构模型
Latent trait model（LTM），潜特质模型（LTM）
Latent variable mean，潜变量均值
Latent variable model，潜变量模型
Least significance difference（LSD）test，最小显著差异(LSD)检验
Levene's test，莱文检验
Likelihood equation，似然方程
Likelihood function，似然函数
Likelihood ratio statistic，似然比统计量
Likelihood ratio test（LRT），似然比检验
Linear growth rate，线性增长率
Linear logit model，线性对数单位模型
Linear model，线性模型
Linear predictor，线性预测值
Linear regression，线性回归
Link function，联系函数
Local dependence model，局部非独立的模型
Local independence，局部独立
Location parameter，位置参数
Logarithm link，对数联系
Logistic distribution，逻辑斯蒂分布
Logistic-ogive model，逻辑斯蒂—S 曲线模型
Logistic regression，逻辑斯蒂回归
Logit link，对数单位联系
Logit model（analysis），对数单位模型（分析）
Log-likelihood function，对数似然函数
Loglinear model（analysis），对数线性模型(分析)
Lognormal distribution，对数正态分布
Lorenz curve，洛伦兹曲线

Mann-Whitney test，曼—惠特尼检验
McNemar test，麦克尼马尔检验
Markov chain model，马尔科夫链模型
Matched studies，配对研究
Maximum likelihood（ML），最大似然
Maximum likelihood estimation（MLE），最大似然估计(MLE)
Mean and dispersion additive model（MADAM），均值和离散相加模型
Mean structure，均值结构
Mean submodel，均值子模型
Measurement error，测量误差
Measurement model，测量模型
MIMIC model，多因多果模型
Mixed-effects model，混合效应模型
Model space，模型空间
Monte Carlo simulation，蒙特卡罗模拟
Multiequation system，多方程系统
Multilevel analysis，多水平分析
Multinomial distribution，多项分布
Multinomial logit link，多项对数单位联系

Risk period，风险期
Risk set，风险集

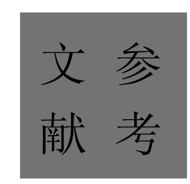
Sampling distribution，抽样分布
Sample size，样本量
Sampling zero，抽到零单元格
Sandwich variance estimator，迫近方差估计量
Scheffé's method，谢弗方法
Schwarz's criterion，施瓦茨准则
Score statistic，得分统计量
Single-step Newton method，单步牛顿方法
Standard (reference) group，标准(参照)组
Standard mortality ratio，标准死亡比
Standard normal distribution，标准正态分布
Standardized rate，标准化比率
Structural equation model，结构方程模型
Structural means model，结构均值模型
Structural model，结构模型
Studentized range distribution，学生化的范围分布
Sufficient statistic，充分统计量
Survival analysis，生存分析
Survival function，生存函数
Systematic component，系统成分
t-distribution，t分布
t-test，t检验
Test statistic，检验统计量
Time varying covariate，时变性协变量
Tolerance distribution，耐受性分布
Treatment group，处理组
Type Ⅰ error，Ⅰ类错误
Tukey's method，图基方法

Uncertainty，不确定性
Underdispersion，离散不足

Variance component，方差成分

Wald statistic，沃尔德统计量
Wilcoxon test，威尔科克森检验

Z-ratio (score)，Z比(得分)
Zero (cell count)，零(单元格数)




Alderson, Arthur S. and François Nielsen. 1999. "Income inequality, development, and dependence: a reconsideration." *American Sociological Review* 64: 606-631.

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Beverly Hills, CA: Sage.

Anderson, Sharon, Ariane Auquier, Walter W. Hauck, David Oakes, Walter Vandaele, and Herbert I. Weisberg. 1980. *Statistical Methods for Comparative Methods: Techniques for Bias Reduction*. New York: Wiley.

Andrews, Donald W. and Ray C. Fair. 1988. "Inference in nonlinear econometric models with structural change." *Review of Economic Studies* 55: 615-640.

Arber, Sara and Helen Cooper. 1999. "Gender differences in health in later life: the new paradox?" *Social Science & Medicine* 48: 61-76.

Arts, Wil and Loek Halman (Eds.). 1999. *New Directions in Quantitative Comparative Sociology*. Boston: Brill

Barndorff-Nielsen, O. E. 1978. *Information and Exponential Families in Statistical Theory*. Winchester: Wiley.

Barndoff-Nielsen, O. E. 1980. "Exponential families." Memoir 5. Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Bartholomew, David J. and Martin Knott. 1999. *Latent Variable Models and Factor Analysis*. Second Edition. Oxford: Oxford University Press.

Becker, Jerry P. (Ed.). 1992. *Report of U.S.-Japan Cross-National Research on Students' Problem Solving Behaviors*. Carbondale, IL: Southern Illinois University.

Blossfeld, H., A. Hamerle, and K. U. Mayer. 1988. *Event History Analysis: Statistical Theory and Applications in Economic and Social Sciences*. Hillsdale, NJ: Erlbaum.

Blossfeld, H. and G. Rohwer. 1995. *Techniques of Event History Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bock, R. D. 1972. "Estimating item parameters and latent ability when responses are scored in two or more nominal categories." *Psychometrika* 37: 29-51.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bornstein, Marc H., O. Maurice Haynes, and Hiroshi Azuma. 1998. "A cross-national study of self-evaluations and attributions in parenting: Argentina, Belgium, France, Israel, Italy, Japan, and the United States." *Developmental Psychology* 34: 662-676.

Breslow, N. E. 1982. "Covariance adjustment of relative-risk estimates in matched studies." *Biometrics* 38: 661-672.

Breslow, N. E. 1996. "Generalized linear models: Checking assumptions and strengthening conclusions." *Statistica Applicata* 8: 23-41.

Browne, Beverly. 1998. "Gender stereotypes in advertising on children's television in the 1990s: a cross-national analysis." *Journal of Advertising* 27: 83-96.

Browne, M. W. 1984. "Asymptotically distribution-free methods for the analysis of covariance structures" *British Journal of Mathematical and Statistical Psychology* 37: 62-83.

Bryk, Anthony S. and Stephen W. Raudenbush. 1996. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.

Buse, A. 1982. "The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note." *The American Statistician* 36: 153-157.

Cai, Jinfa. 1995. "A cognitive analysis of U. S. and Chinese students' mathematical performance on tasks involving computation, simple problem solving, and complex problem solving." *Journal for Research in Mathematics Education Monograph Series*, No. 7. Reston, VA: National Council of Teachers of Mathematics.

Choi, Namkee G. 1997. "Racial differences in retirement income: the roles of public and private income sources." *Journal of Aging & Social Policy* 9: 21-42.

Chow, Gregory C. 1960. "Tests of equality between sets of coefficients in two linear regressions." *Econometrica* 28: 591-605.

Clogg, Clifford C. 1978. "Adjustment of rates using multiplicative models." *Demography* 15: 523-539.

Clogg, Clifford C. 1981. "New developments in latent structure analysis." pp. 215-246 in *Factor Analysis and Measurement in Sociological Research*, edited by D. J. Jackson and E. F. Borgotta. Beverly Hills, CA: Sage.

Clogg, Clifford C. and Scott R. Eliason. 1987. "Some common problems in log-linear analysis." *Sociological Methods Research* 16: 8- 44.

Clogg, Clifford C. and Scott R. Eliason. 1988. "A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons." *American Sociological Review* 53: 267-283.

Clogg, Clifford C. and Leo A. Goodman. 1984. "Latent structure analysis of a set of multidimensional contingency tables." *Journal of American Statistical Association* 79: 762-771.

Clogg, Clifford C. and Leo A. Goodman. 1985. "Simultaneous latent structure analysis in several groups." pp. 81-110 in *Sociological Methodology 1985*, edited by Nancy B. Tuma. San Francisco: Jossey-Bass.

Clogg, Clifford C. and James W. Shockey. 1988. "Multivariate analysis of discrete data." In *Handbook of Multivariate Experimental Psychology*, edited by J. R. Nesselroade and R. B. Cattell. New York: Plenum.

Clogg, Clifford C., James W. Shockey, and Scott R. Eliason. 1990. "A general statistical framework for adjustment of rates." *Sociological Methods & Research* 19: 156-195.

Courgeau, D. and E. Lelievre. 1992. *Event History Analysis in Demography*. Oxford: Clarendon.

Cox, D. R. 1972. "Regression models and life tables." *Journal of the Royal Statistical Society B* 34: 187-203.

Cox, D. R. 1975. "Partial Likelihood." *Biometrika* 62: 269-276.

Cox, D. R. and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman and Hall.

Cressie, N. and P. W. Holland. "Characterizing the manifest probabilities of latent trait models." *Psychometrika* 48: 129-141.

Croon. M. 1990. "Latent class analysis with ordered classes." *British Journal of the Mathematical Statistical Society* 43: 171-192.

Das Gupta, Biman Kumar. 1994. "The comparative method in anthropology." *Current Anthropology*

35: 558-559.

Das Gupta, Prithwis. 1991. "Decomposition of the difference between two rates and its consistency when more than two populations are involved." *Mathematical Population Studies* 3:125.

Demo, D. H. and K. D. Parker. 1987. "Academic achievement and self-esteem among black and white college students." *Journal of Social Psychology* 127:345-355.

Descartes, René. 1963. *OEuvres Philosophiques*, Vol. 1, edited by Ferdinand Alqui, Paris: Garnier.

Dey, Dipak, Sunjit K. Ghosh, and Bani K. Mallick. 2000. *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker.

Dobson, Annette J. 1990. *An Introduction to Generalized Linear Models*. London: Chapman and Hall.

Echevarria, Cristina and Antonio Merlo. 1999. "Gender differences in education in a dynamic household bargaining model." *International Economic Review* 40:265-286.

Efron, Bradley. and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Ehrhardt-Martinez, Karen. 1998. "Social determinants of deforestation in developing countries: a cross-national study." *Social Forces* 77:567-586.

Ellina, Maro and Will H. Moore. 1990. "Discrimination and political violence: a cross-national study with two time periods." *Western Political Quarterly* 43: 267-278.

Fisher, R. A. 1932. *Statistical Methods for Research Workers*. Fourth Edition. Edinburgh: Oliver and Boyd.

Fisher, R. A. 1935. *The Design of Experiments*. Frist Edition. Edinburgh and London: Oliver and Boyd.

Fleiss, Joseph L. 1981. *Statistical Methods for Rates and Proportions*. Second Edition. New York: Wiley.

Fleiss, Joseph L. 1986. *The Design and Analysis of Clinical Experiments*. New York: Wiley.

Gershuny, Jonathan. 2000. *Changing Times: Work and Leisure in Postindustrial Society*. Oxford: Oxford University Press.

Goethe, Johann Wolfgang von. 1982. *Werke: Hamburger Ausgabe*, Vol. 13. Munich: DTV.

Goldstein, H. 1995. *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H., J. Rasbash, W. Browne, G. Woodhouse, and M. Poulain. 2000. "Multilevel models in the study of dynamic household structures." *European Journal of Population* 16:373-387.

Good, Phillip. 2000. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Second Edition. New York: Springer.

Goodman, Leo A. 1973. "The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach." *Biometrika* 60:179-192.

Greene, William H. 1990. *Econometric Analysis*. New York: Macmillan.

Haberman, Shelby J. 1979. *Analysis of Qualitative Data*, Vol. 2: *New Developments*. New York: Academic.

Hagenaars, Jacques A. 1988. "Latent structure models with direct effects between indicators: local dependence models." *Sociological Methods Research* 16:379-405.

Hagenaars, Jacques A. 1993. *Loglinear Models with Latent Variables*. Newbury Park, CA: Sage.

Handcock, Mark S. and Martina Morris. 1998. "Relative distribution methods." pp. 53-97 in *Sociological Methodology 1998*, edited by Adrian E. Raftery. Boston: Blackwell.

Handcock, Mark S. and Martina Morris. 1999. *Relative Distribution Methods in the Social Sciences*. New York: Springer-Verlag.

Hastie, T. J. and R. J. Tibshirani. 1990. *Generalized Additive Models*. London: Chapman and Hall.

Heinen, Ton. 1993. *Discrete Latent Variable Models*. Tilburg: Tilburg University Press.

Henriques, Gregg R. and Lawrence G Calhoun. 1999. "Gender and ethnic differences in the relationship between body esteem and self-esteem." *The Journal of Psychology* 133:357-368.

Higgins, Joan. 1981. *States of Welfare: Comparative Analysis in Social Policy*. Oxford: Blackwell and Martin Robertson.

Hill, P. W. and H. Goldstein. 1998. "Multilevel modelling of educational data with cross classification and missing identification of units." *Journal of Educational Behavioural Statistics* 23: 117-128.

Honda, Yuzo and Kazuhiro Ohtani. 1986. "Modified Wald tests in tests of equality between sets of coefficients in two linear regressions under heteroscedasticity." *The Manchester School of Economic and Social Studies* 54: 208-218.

Hox, J. J. 1994. *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties.

Hupkens, Christianne L. H., Ronald A. Knibbe, and Anneke H. van Otterloo. 1998. "Class differences in the food rules mothers impose upon their children: a cross-national study (Netherlands, Belgium, and Germany)." *Social Science & Medicine* 47:1331-1339.

Ishida, Hiroshi, Walter Muller, and John M Ridge. 1995. "Class origin, class destination, and education: a cross-national study of ten industrial nations." *American Journal of Sociology* 101: 145-193.

Johnson, Monica Kirkpatrick and Margaret Mooney Marini. 1998. "Bridging the racial divide in the United States: the effects of gender." *Social Psychology Quarterly* 61: 247-258.

Kalbfleisch, J. D. and R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Kass, Robert E. and Adrian E. Raftery. 1995. "Bayes factor." *Journal of the American Statistical Association* 90: 773-795.

Kenworthy, Lane. 1999. "Do social-welfare policies reduce poverty? A cross-national assessment." *Social Forces* 77: 1119-1139.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Cambridge: Cambridge University Press.

Kitagawa, E. M. 1955. "Components of a difference between two rates." *Journal of the American Statistical Association* 50:1168-1194.

Kluegel, James R., David S. Mason, and Bernd Wegener (Eds.). 1995. *Social Justice and Political Change: Public Opinion in Capitalist and Post-Communist States*. Hawthorne, NY: Aldine de Gruyter.

Kohn, Melvin L. (Ed.). 1989. *Cross-National Research in Sociology*. American Sociological Association Presidential Series. Newbury Park, CA: Sage.

Krall, J. M., V. A. Uthoff, V. A. and J. B. Harley. 1975. "A step-up procedure for selecting variables associated with survival." *Biometrics* 31: 49-57.

Kramer, C. Y. 1956. "Extension of multiple range test to group means with unequal number of replications." *Biometrics* 12: 307-310.

Kreft, I. G. and J. de Leeuw. 1998. *Introducing Multilevel Modeling*. Thousand Oaks, CA: Sage.

Lacour, Claudia Brodsky. 1995. "Grounds of comparison." *Would Literature Today* 69: 271-274.

Lacy, Michael G. 1997. "Efficiently studying rare events: case-control methods for sociologists." *Sociological Perspectives* 40: 129-154.

Lancaster, T. 1990. *The Economic Analysis of Transition Data*. Cambridge: Cambridge University Press.

Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data*. New York: Wiley.

Lazarsfeld, Paul F. and Neil W. Henry. 1968. *Latent Structure Analysis*. Boston: Houghton Mifflin.

Lesaffre, E. and B. Spiessens. 2001. "On the effect of the number of quadrature points in a logistic random-effects model: An Example." *Journal of the Royal Statistical Society Series C—Applied Statistics* 50(Part 3):325-335.

Levene, H. 1960. "Robust tests for equality of variances." pp. 278-292 in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin et al. Stanford, CA: Stanford University, Press.

Liao, F. 1989a. "A flexible approach for the decomposition of rate differences." *Demography* 26: 717-726.

Liao, F. 1989b. " Fertility differentials of religious and ethnic groups in the United States: a fuzzy group membership approach. " Ph. D. Dissertation. University of North Carolina at Chapel Hill.

Liao, F. 1989c. "Estimating fuzzy set membership coefficients with log-multiplicative association models: the case of contraceptiveness." *Mathematical Population Studies* 1:357-376.

Liao, T. F. 2002. "Bayesian model comparison in generalized linear models across multiple grorps." *Computational Statistics and Data Analysis* 39.

Lloyd, Kim M. and Scott J. South. 1996. "Contextual influences on young men's transition to first marriage." *Social Forces* 74: 1097-1119.

Longford, N. T. 1993. *Random Coefficient Models*. New York: Oxford University Press.

Lorenz, Max O. 1905. "Methods of measuring the concentration of wealth." *Journal of the American Statistical Association* 9: 209-219.

Mabbett, Deborah and Helen Bolderson. 1999. "Theories and Methods in Comparative Social Policy." pp. 34-56 in *Comparative Social Policy: Concepts, Theories and Methods*, edited by Jochen Clasen. Oxford: Blackwell.

Manton, K. G., K. Liu, and E. S. Cornelius. 1985. "An analysis of the heterogeneity of U. S. nursing home patients." *Journal of Gerontology* 40: 34-46.

Manton, K. G., E. Stallard, M. A. Woodbury, H. D. Tolley, and A. I. Yashin. 1987. "Grade-of-membership techniques for studying complex event history processes with unobserved covariates." pp. 309-346. *Sociological Methodology 1987*, edited by C. C. Clogg. Washington, DC: American Sociological Association.

Martin, William G. and Mark Beittel. 1998. "Toward a global sociology? Evaluating current conceptions, methods, and practices." *Sociological Quarterly* 39:139-161.

McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models*. Second Edition. London: Chapman and Hall.

McCulloch, C. E. 1997. "Maximum likelihood algorithms for generalized linear mixed models." *Journal of the American Statistical Association* 92: 162-170.

McDowell, John M., Larry D. Singell Jr., and James P. Ziliak. 1999. "Cracks in the glass ceiling: gender and promotion in the economics profession." *American Economic Review* 89: 392-396.

McMichael, Philip. 1990. " Incorporating comparison within a world-historical perspective: an alternative comparative method." *American Sociological Review* 55: 385-397.

McNemar, Q. 1947. "Note on the sampling error of the difference between correlated proportions or percentages." *Psychometrika* 12: 153-157.

Mennen, Ferol E. 1994. "Sexual abuse in Latina girls: their functioning and a comparison with white and African American girls." *Hispanic Journal of Behavioral Sciences* 16: 475-486.

Miller, John William. 1981. *The Philosophy of History with Reflections and Aphorisms*. New York: Norton.

Mitchell, W. J. T. 1996. "Why comparisons are odious." *World Literature Today* 70: 321-324.

Moustaki, I and M. Knott. 2000. "Generalized latent trait models." *Pychometrika* 65: 391-411.

Mueller, Ralph O. 1996. *Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS*. New York: Springer.

Nelder, J. A. 1998. "A large class of models derived from generalized linear models." *Statistics in Medicine* 17: 2747-2753.

Nelder, J. A. and R. W. M. Wedderburn. 1972. "Generalized linear models." *Journal of the Royal Statistical Society A* 135: 370-384.

Nowak, Stefan. 1989. "Comparative studies and social theory." pp. 34-56 in *Cross-National Research in Sociology*, edited by Melvin L. Kohn. American Sociological Association Presidential Series. Newbury Park, CA: Sage.

Otterbein, Keith F. 1994. "The comparative method in anthropology." *Current Anthropology* 35: 559-560.

Perrucci, Carolyn C., Robert Perrucci, and Dena B. Targ. 1997. "Gender differences in the economic, psychological and social effects of plant closings in an expanding economy." *Social Science Journal* 34: 217-233.

Peters, B. Guy. 1998. *Comparative Politics: Theory and Methods*. New York: New York University Press.

Raftery, Adrian. 1995. "Bayesian model selection in social research." pp. 111-163 in *Sociological Methodology 1995*, edited by Peter V. Marsden. Cambridge, MA: Blackwell.

Raftery, Adrian. 1996. "Approximate Bayes factors and accounting for model uncertainty in generalised linear models." *Biometrika* 83: 251-266.

Rao, J. N. K. and A. J. Scott. 1992. "A simpie method for the analysis of clustered binary data." *Biometrics* 48: 577-586.

Rao, J. N. K. and A. J. Scott. 1999. "A simple method for analysing overdispersion in clustered Poisson data." *Statistics in Medicine* 18: 1373-1386.

Ridley, Mark. 1994. "The comparative method in anthropology." *Current Anthropology* 35: 560-561.

Rigby R. A. and M. D. Stasinopoulos 1995. "Mean and dispersion additive models: applications and diagnostics." pp. 249-256 in *Statistical Modelling: Proceedings of the 10th International Workshop on Statistical Modelling*, edited by G. U. H. Seeber, B. J. Francis, R. Hatzinger, and G. Steckel-Berger. New York: Springer-Verlag.

Rigby R. A. and M. D. Stasinopoulos 1996. "A semi-parametric additive model for variance heterogeneity." *Statistics and Computing* 6: 57-65.

Sahai, Hardeo and Mohammed I. Ageel. 2000. *The Analysis of Variance: Fixed, Random, and Mixed Models*. Boston: Birkhauser.

Sapiro, Virginia and Pamela Johnston Conover. 1997. "The variable gender basis of electoral politics: gender and context in the 1992 US election." *British Journal of Political Science* 27: 497-523.

Schall, Robert. 1991. "Estimation in generalized linear models with random effects." *Biometrika* 78: 719-727.

Scheffé, H. 1953. "A method for judging all contrasts in the analysis of variance." *Biometrika* 40: 87-104.

Scheffé, H. 1959. *The Analysis of Variance*. New York: Wiley.

Schweizer, Thomas. 1994. "The comparative method in anthropology." *Current Anthropology* 35: 561.

Silver, Edward A., S. S. Leung, and Jinfa Cai. 1995. "Generating multiple solutions for a problem: a comparison of the responses of U.S. and Japanese students." *Educational Studtes in Mathematics* 28: 35-54.

Smith, Herbert L. 1997. "Matching with multiple controls to estimate treatment effects in observational studies" pp. 325-353 in *Sociological Methodology 1997*, edited by Adrian E. Raftery. Boston: Blackwell.

Smith, Philip J. and Daniel F. Heitjan. 1993. "Testing and adjusting for departures from nominal dispersion in generalized linear models." *Applied Statistics*. 42: 31-41.

Smyth, G. K. 1989. "Generalized linear models with varying dispersion." *Journal of Royal Statistical Society B*, 51: 47 -60.

Smyth, G. K. and A. P. Verbyla, 1996. "A conditional approach to residual maximum likelihood estimation in generalized linear models." *Journal of Royal Statistical Society B*, 58: 565-572.

Smyth, G. K. and A. P. Verbyla, 1999. "Adjusted likelihood methods for modelling dispersion in generalized linear models." *Environmetrics* 10: 695-709.

Snijders, T. and R. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Sörbom, D. 1981. "Structural equation models with structured means." In *Systems under Indirect Observation: Causality, Structure and Prediction*, Vol. 1, edited by K. G. Jreskog and H. Wold. Amsterdam: North-Holland Publishing Co.

Sullivan, J. L. and J. E. Transue. 1999. "The psychological underpinnings of democracy: a selective review of research on political tolerance, interpersonal trust, and social capital." *Annual Review of Psychology* 50: 625-650.

Telles, Edward E. and Nelson Lim. 1998. "Does it matter who answers the race question? Racial classification and income inequality in Brazil." *Demography* 35: 465-474.

Tukey, John W. 1953. "The problem of multiple comparisons" (Mimeograph, 396 pages). Princeton, NJ: Department of Mathematics, Princeton University.

Tuma, Nancy B. and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods*. New York: Academic.

Vermunt, Jeroen K. 1997. *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage.

Voland, Eckart. 1994. "The comparative method in anthropology." *Current Anthropology*: 35: 561-562.

Wald, Abraham. 1943. "Tests of statistical hypothesis concerning several parameters when the number of observations is large." *Transactions of the American Mathematical Society* 54: 426-482.

Wesolowski, Wlodzimierz. 1989. "Legitimate domination in comparative-historical perspective: the case of legal domination." pp. 34-56 in *Corss-National Research in Sociology*, edited by Melvin L. Kohn. American Sociological Association Presidential Series. Newbury Park, CA: Sage.

Westholm, Anders and Richard G. Niemi. 1992. "Political institutions and political socialization: a cross-national study." *Comparative Politics* 25: 25-41.

Wonnacott, Ronald J. and Thomas H. Wonnacott. 1979. *Econometrics*. Second Edition. New York: Wiley.

Wright, Erik Olin, Janeen Baxter, and Gunn Elizabeth Birkelund. 1995. "The gender gap in workplace authority: a cross-national study." *American Sociological Review* 60: 407-435.

Yamaguchi, Kazuo. 1991. *Event History Analysis*. Newbury Park, CA: Sage.